**Katholieke Universiteit Leuven**
**Group Biomedical Sciences**
**Faculty of Pharmaceutical Sciences**
**Rega Institute for Medicinal Chemistry**

# Replacing the pyrophosphate leaving group in the enzymatic polymerization of nucleotides to DNA

Anne GIRAUT

Doctoral thesis in Pharmaceutical Sciences

Leuven, 2010

**Katholieke Universiteit Leuven**
**Group Biomedical Sciences**
**Faculty of Pharmaceutical Sciences**
**Rega Institute for Medicinal Chemistry**

# Replacing the pyrophosphate leaving group in the enzymatic polymerization of nucleotides to DNA

Anne GIRAUT

Jury:

Promoter: Prof. Dr. Piet Herdewijn
Co-promoter: Prof. Dr. Arthur Van Aerschot
Chair: Prof. Dr. Ann Van Schepdael
Secretary: Pr. Dr. Peter de Witte
Jury members: Prof. Dr. Peter de Witte
Prof. Dr. Johan Robben
Dr. Dieter Heindl

Leuven, 06.12.2010
Doctoral thesis in Pharmaceutical Sciences

## Professional career Anne Giraut

Heindl D., Kessler D., Schube A., Thuer W., **Giraut A.,** (2008) Easy method for the synthesis of labeled oligonucleotides. *Nucleic Acid Symp. Ser. (Oxf.),* **52**, 405-406.

**Giraut A.**, Dyubankova N., Song X.-P., Herdewijn, P. (2009) Phosphodiester substrates for incorporation of nucleotides in DNA using HIV-1 reverse transcriptase. *Chembiochem* **10,** 2246-2252.

Zlatev I., **Giraut A.**, Morvan F., Herdewijn P., and Vasseur J. J. (2009) δ-Di-carboxybutyl phosphoramidate of 2 '-deoxycytidine-5 '-monophosphate as substrate for DNA polymerization by HIV-1 reverse transcriptase. *Bioorganic & Medicinal Chemistry* **17,** 7008-7014.

**Giraut A.**, Song X.-P., Froeyen M., Marlière P., Herdewijn P. (2010) Iminodiaceic acid-phosphoramidates as metabolic prototypes for diversifying nucleic acid synthesis *in vivo*. *Nucleic Acid Research* **38**, 2541-2550.

**Giraut A.**, Herdewijn P. (2010) Influence of the linkage between leaving group and nucleoside on substrate efficiency for incorporation in DNA catalyzed by reverse transcriptase. *Chembiochem* **11**, 1399-1403.

**Giraut A.**, Song, X.-P., Herdewijn, P., (2010) Mimicking the pyrophosphate leaving group in the enzymatic synthesis of DNA by polymerases. *International Round Table on Nucleosides, Nucleotides and Nucleic Acids,* Lyon, France.

# ACKNOWLEDGEMENTS

*I would like to thank the EC (Orthosome Strep Project) and the KULeuven (GOA grant) for financial support.*

*This thesis is the result of four years of work during which I have been accompanied and supported by many people. I would like here to express my gratitude to all of them. Dieter Heindl is the reason I started this PhD in the first place and for this, for his advice and encouragement, I thank him enormously.*

*I would like to thank my supervisor Prof. Dr. Piet Herdewijn for giving me the opportunity to do a doctoral thesis as part of an exciting project, for introducing me to the study of polymerases and to the pioneering field of synthetic biology.*

*I also thank my co-promotor Prof. Dr. Arthur Van Aerschot for his friendly supervision and precious advice, and for his patience in helping me through the (somewhat complicated) procedures of the doctoral school.*

*I would like to express my gratitude to all the jury members for accepting to be a member of the evaluating committee and for reviewing my manuscript.*

*I thank Dr. Montse Terrazas for her guidance during my first steps in the lab, Luc Baudemprez for recording the 500 MHz NMR spectra, Prof. Dr. Jef Rozenski for mass spectrometry and LCMS measurement and advice, Guy Schepers for his joyful company and for letting me borrow his lab equipment again and again, Wim Spreutels for his help in biological experiments and friendly lab sharing. I am grateful to all the referees of the articles, as well as Drs. Olga Adelfinskaya and Susan Cure for carefully reading and correcting my manuscripts. I am also grateful to Chantal Biernaux, Mia Vanthienen for all the editorial and practical help, and to Dominique Brabants, Pierre Fiten, Inge Aerts and Christiane Callebaut for their secretarial and practical help. I had the pleasure to work with Marleen Renders, Cécile Gasse, Elisabetta Groaz, Céline Crauste, Camille Bouillon, Sophie Duraffour and Dimitri Topalis. I want to thank them for their valuable advice, helpful scientific discussions and support during the realisation steps of this manuscript. I would also like to thank all my colleagues of the laboratory for Medicinal Chemistry not mentioned above; I am grateful to them for their help and for the friendly atmosphere reigning in the lab.*

*I thank my sister, for everything; Audrey, Sophie, Cécile for their unconditional support and friendship, and Dries for his love and patience. I thank my parents, les Caru, all my friends and family for encouraging me during my unexpectedly prolonged studies and for their love and support.*

# ABSTRACT

The polymerization of nucleotides into nucleic acids is catalyzed by polymerases and is the supporting process for the survival and reproduction of all living organisms. In recent decennia analogues of nucleoside triphosphates ((d)NTPs), precursors of nucleic acids, have been synthesized in order to study the polymerization mechanism. Hereto, synthetic modifications can be introduced on different parts of the precursors. For instance, polymerase recognition of (d)NTP analogues where pyrophosphate has been replaced by various candidate leaving groups informs us about the mode of action of DNA polymerization enzymes. In addition, efficient leaving groups differing form pyrophosphate could be used for the development of new antivirals or serve for the orthogonalization of nucleic acid synthesis. More importantly, such leaving groups could support an orthogonal genetic system that cannot interfere with the natural cellular system.

In the second chapter various chemical groups are investigated as mimics for pyrophosphate in the polymerization of DNA by HIV-1 reverse transcriptase. Diverse aliphatic and aromatic structures carrying one or two carboxylic acid functions are able to replace pyrophosphate during the nucleic acid synthesis process, albeit at the expense of DNA extension efficiency. During the nucleotide incorporation step phosphoramidate analogues seem to be less easily cleaved by the polymerase than their phosphodiester counterparts. Structures carrying only one carboxylic acid function can mimic pyrophosphate only when they are connected to the nucleoside *via* a phosphodiester linkage.

In the third chapter the influence of the nature of this linkage is further investigated. Replacing the phosphoramidate linkage with a phosphodiester leads to better single nucleotide incorporation results. Longer DNA extension seems to be invariably impaired by the presence of a modified leaving group. We demonstrate that the nature of the linkage is not responsible for DNA synthesis pausing.

The fourth chapter introduces a different leaving group mimic, iminodiacetic acid (IDA). IDA-deoxyadenosine monophosphate is a better substrate for single nucleotide incorporation and primer elongation than previous analogues. The reasons for DNA synthesis pausing are further explored. Computer modelling of ground state interactions during the chemical step of the process seems to support the hypothesis that IDA-dAMP binds to the polymerase active site in a similar way as compared to the natural substrate. Metabolic accessibility of this alternative DNA building block starting from naturally occurring

cyclotriphosphate is explored and shows that such DNA building blocks could potentially be biosynthesized. Directed evolution of polymerases is needed to develop enzymes capable of efficiently and exclusively using IDA-dNMPs as substrates.

Metabolic availability of IDA-dNMPs is further explored in chapter five. IDA substituted adenosine triphosphate is synthesized and the ability of a human kinase to use it as a phosphate donor is tested. Deoxycytidine kinase cannot catalyze the transfer of IDA-phosphoryl onto its natural substrates, showing the need for bacterial or engineered enzymes endowed with broader substrate specificity. Finally a bacterial enzyme capable of degrading the released leaving group is identified, bringing us a step closer to an orthogonal nucleic acid machinery.

We could conclude that the pyrophosphate moiety of nucleotide triphosphates can be replaced by various leaving groups during the enzymatic synthesis of DNA. The efficiency of this synthesis depends on the structure of the group and on the nature of the linkage between the leaving group and the nucleotide to be incorporated. Further enzymatic studies are needed to understand the mechanisms of recognition and processing of polymerases. Directed evolution of polymerases is the following and necessary step in the development of polymerases that can use IDA-deoxynucleoside monophosphates as substrates in an efficient and exclusive manner.

# SAMENVATTING

De polymerisatie van nucleotiden tot nucleïnezuur wordt gekatalyseerd door polymerasen en dit proces is essentieel voor de overleving en reproductie van alle levende organismen. De laatste decennia werden tal van synthetische analogen van nucleoside-trifosfaten getest als substraat voor verschillende polymerasen met als doel het polymerisatiemechanisme te bestuderen. Chemische synthesemethoden maken het ons mogelijk om elk gedeelte van een nucleoside trifosfaat te modificeren en zo kan het pyrofosfaatgedeelte van een nucleoside trifosfaat vervangen worden door een andere uittredende groep. Een dergelijke studie kan ons veel leren over het mechanisme van enzymatische DNA synthese. Daarnaast kan deze technologie ook toegepast worden voor de ontwikkeling van een nieuw type van antivirale middelen of voor nucleïnezuursynthese. Tenslotte zijn zulke gemodificeerde nucleotiden ook geschikt om nieuwe genetische systemen te ontwikkelen die orthogonaal zijn aan het natuurlijke cellulaire systeem en er dus niet mee kunnen interfereren.

In het eerste experimentele hoofdstuk worden verschillende chemische groepen onderzocht als 'mimics' voor pyrofosfaat in de polymerisatie van dAMP tot DNA door HIV-1 reverse transcriptase (RT). Diverse alifatische of aromatische groepen die één of twee carbonzuurfuncties dragen, kunnen pyrofosfaat vervangen in het nucleïnezuursyntheseproces, ook al is de efficiëntie lager dan bij het gebruik van natuurlijke trifosfaten. Het inbouwen van nucleotiden lijkt minder efficiënt te verlopen met een fosforamidaatanaloog dan met zijn fosfodiëstertegenhanger. Structuren met slechts één carbonzuurfunctie werken enkel als pyrofosfaat-'mimics' als de binding die doorbroken wordt tijdens de polymerisatie een fosfodiëster is.

In het tweede experimentele hoofdstuk wordt de invloed van de aard van de binding tussen de uittredende groep en het nucleotide verder onderzocht. Het vervangen van een fosforamidaatbinding door een fosfodiësterbinding vertaalt zich in een efficiëntere inbouw van het nucleotide in DNA. De synthese van langere DNA-ketens waarbij gebruik gemaakt wordt van deze gemodificeerde uittredende groepen wordt echter moeilijk. Een interessante waarneming is dat de ketenverlenging niet afhankelijk is van de aard van de binding tussen de fosfaatgroep en de uittredende groep.

Het derde experimentele hoofdstuk beschrijft de synthese en eigenschappen van een nieuwe uittredende groep, het iminodiazijnzuur (IDA). IDA-deoxyadenosine-monofosfaat is een beter substraat voor de enzymatische synthese van DNA (verlenging van DNA-primers

met één of meerdere nucleotiden) dan de vroegere onderzochte analogen. De reden voor de vroegtijdige onderbreking van DNA-synthese werd verder onderzocht. Computermodelling van de binding van IDA-dAMP in de actieve plaats van HIV-1 RT lijkt de hypothese te ondersteunen dat IDA-dAMP zich aan het enzym bindt op een gelijkaardige manier als het natuurlijke substraat. Daarna werd de metabole toegankelijkheid van deze alternatieve DNA-bouwsteen bestudeerd. We hebben aanwijzigingen dat natuurlijk voorkomend cyclisch trifosfaat een belangrijke rol kan spelen in zijn biosynthese. Een verdere stap in het onderzoek is de gestuurde evolutie van polymerasen die efficiënt en exclusief IDA-dNMPs als substraat kunnen gebruiken.

De metabole toegankelijkheid van IDA-dNMPs werd verder onderzocht in het experimentele hoofdstuk vier. Hierbij werd IDA-gesubstitueerde adenosine-trifosfaat chemisch gesynthetiseerd. Het vermogen van een eoxycytidine kinase van menselijke oorsprong om IDA-ATP als fosfaatbron te gebruiken werd onderzocht. Helaas kan dit deoxycytidine-kinase de overdracht van IDA-monofosfaat naar hun natuurlijke substraten (dA of dC) niet katalyseren. Verder onderzoek dient uit te maken of bacteriële (of gemodificeerde) enzymen een bredere substraatspecificiteit hebben. Uiteindelijk werd ook één bacterieel enzym geïdentificeerd dat wel in staat bleek de afbraak van IDA te katalyseren waardoor het polymerisatieproces irreversibel wordt.

We kunnen besluiten dat het pyrofosfaatgedeelte van nucleoside-trifosfaten kan vervangen worden door verschillende andere uittredende groepen tijdens het enzymatisch syntheseproces van DNA. De efficiëntie van deze synthese is afhankelijk van de structuur en van de aard van de binding tussen de uittredende groep en het nucleotide dat in het DNA wordt ingebouwd. Dit onderzoek is slechts een begin en nog vele functionele enzymstudies zijn nodig om dit proces van herkenning en verwerking bij polymerasen te begrijpen. Gestuurde evolutie van polymerasen is nodig om enzymen te ontwikkelen die op een efficiënte manier uitsluitend IDA-deoxynucleoside-monofosfaten kunnen gebruiken als substraten.

# LIST OF ABREVIATIONS

| | |
|---|---|
| *(i)*PA | (iso)phtalic acid |
| A | adenosine |
| AMV | alfalfa mosaic virus |
| AS | antisense |
| (L-) Asp | (L-) aspartic acid |
| $Bu_3N$ | tri-*n*-butylamine |
| C | cytidine |
| CDI | N, N'-carbonyldiimidazole |
| CeNA | cyclohexenyl nucleic acid |
| CSR | compartmentalized self replication |
| d | deoxy- |
| DCC | dicyclohexylcarbodiimide |
| dCK | deoxycytidine kinase |
| dd | dideoxy- |
| DIPA | diisopropylamine |
| DIPEA | diisopropylethylamine |
| DMF | dimethylformamide |
| DMSO | dimethylsulfoxide |
| DMTr | dimethoxytrityl |
| DNA | deoxyribonucleic acid |
| dNDP | deoxynucleoside diphosphate |
| dNMP | deoxynucleoside monophosphate |
| dNTP | deoxynucleoside triphosphate |
| ds | double strand(ed) |
| EDTA | ethylene diamine tetra acetic acid |
| ESI | electron spray ionization |
| $Et_3N$ | triethylamine |
| G | guanosine |
| GA | glycolic acid |
| HAART | highly active antiretroviral therapies |
| (L-) His | (L-) histidine |
| HIV-1 | human immunodeficiency virus 1 |
| HNA | hexitol nucleic acid |
| HRMS | high resolution mass spectrometry |
| HSV | herpes simplex virus |
| HTS | high troughput screening |
| I | inosine |

| | |
|---|---|
| IDA | iminodiacetic acid |
| IE | ion exchange |
| ILA | imidazole lactic acid |
| Im | imidazole |
| kDa | kiloDaltons |
| $K_M$ | Michaelis-Menten constant |
| mRNA | messenger RNA |
| MA | malic acid |
| MLV | murine leukemia virus |
| ms | mass spectrometry |
| NAD | Nicotine amide dinucleotide |
| NEM | N-ethyl morpholine |
| NMR | nuclear magnetic resonance |
| NRTI | Nucleotide-like reverse transcriptase inhibitors |
| ON | oligonucleotide |
| P:T | primer template hybridized duplex |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| PMEA | phosphonylmethoxyethyl adenine |
| PMO | phosphorodiamidate morpholino oligonucleotide |
| $PPh_3$ | triphenylphosphine |
| PPi | pyrophosphate |
| PS | phosphorothioate |
| RNA | ribonucleic acid |
| RP | reverse phase |
| RT | reverse transcriptase |
| r.t. | room temperature |
| ss | single strand(ed) |
| T | thymidine |
| Taq polymerase | DNA polymerase of *Thermus aquaticus bacterium* |
| TBE | tris-borate EDTA |
| TEAA | triethylammonium acetate |
| TEAB | triethylammonium bicarbonate |
| THF | tetrahydrofurane |
| TLC | thin layer chromatography |
| U | uridine |
| UV | ultraviolet |
| $V_{max}$ | maximum velocity of incorporation |
| XNA | xenonucleic acid |

# TABLE OF CONTENTS

# Chapter I – Introduction

## *I.1 Nucleic acids*

The term 'nucleic acid' refers to informational polymers containing genetic material. It includes deoxyribonucleic acids (DNA) and ribonucleic acids (RNA). The monomer units, respectively deoxynucleotides (for DNA) or nucleotides (for RNA), are composed of a β-D-2′-deoxyribose (DNA) or β-D-ribose (RNA) sugar covalently bound to a) a phosphate moiety at the 5′-C and b) to a heterocyclic base bound at the 1′-position through a β-glycosidic C1-N linkage (Figure I.1). The term '(deoxy)nucleoside' is used to refer to a (deoxy)nucleotide deprived of the phosphate group. The heterocyclic bases have either a purine (adenine or guanine in the case of adenosine and guanosine nucleosides, rspectively) or a pyrimidine (cytosine, thymine and uracile in the case of cytosine, thymidine and uridine nucleosides, respectively) structure.

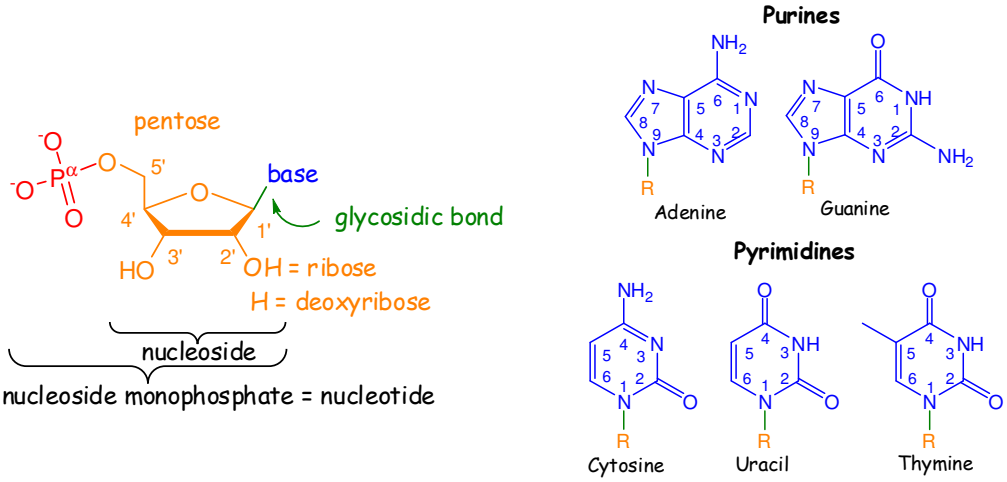



**Figure I.1:** Left: general structure of (deox)nucleotides and conventional numbering of the sugar backbone; Right: chemical structures of the five heterocyclic bases of DNA/RNA: adenine, guanine, cytosine, uracil and thymine and conventional numbering of the cyclic atoms.

A DNA strand is made of nucleotides connected *via* 5′-3′ phosphodiester bonds. The resulting sugar-phosphate chain is referred to as the backbone of nucleic acids. In living organisms, RNA occurs as a single-stranded molecule, whereas DNA mostly occurs as a

double-stranded helix. This double helix is stabilized by hydrogen-bonding interactions between the bases of opposite strands. In the classical 'complementary base pairing' system unravelled by Watson and Crick in 1953, adenine pairs with thymine *via* two hydrogen-bonds, while guanine pairs with cytosine *via* three hydrogen-bonds (Figure I.2) (1).

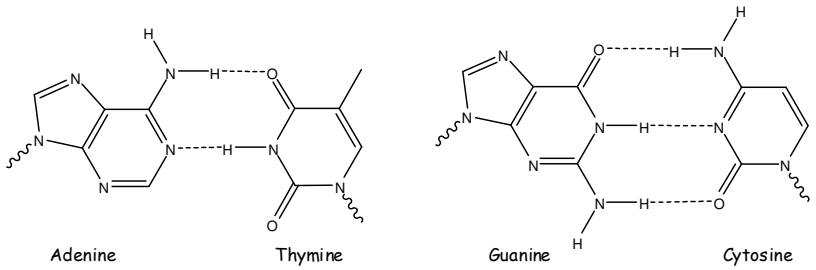



**Figure I.2:** Classical Watson-Crick complementary base pairs.

In a nucleic acid chain the five-membered sugar ring of a nucleotide adopts a non-planar conformation resembling a half-chair (Figure I.3). As a consequence, and depending on nucleotide sequence, osmotic conditions or replication stage, a DNA helix can adopt three conformations that differ in geometry and dimensions (Figure I.4) (2).

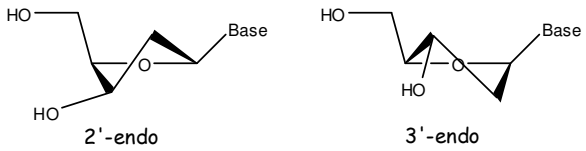



**Figure I.3:** Spatial representation of possible conformations adopted by deoxyribonucleosides in nucleic acids.

Due to the 2′-endo ribose conformation, DNA usually forms a B-helix. The A-type helix conformation is adopted by dsRNA and DNA/RNA hybrids and is a consequence of the 3′-*endo* conformation found in RNA. Both forms of helix are right-handed and anti-parallel (opposite 'reading' direction) and present a minor and a major groove to the environment. In the A double helix the major groove is very deep and narrow and the minor groove is wide and shallow. In the B-double helix both grooves have a similar depth, but their widths differ as the major groove is wide while the minor groove is narrow. A third type of helix called Z-DNA is less common and can arise in the case of alternating purine-pyrimidine sequences. The purine nucleotides then adopt a 2′-*endo* conformation, while the pyrimidine nucleotides adopt a 3′-*endo* conformation, resulting in a left-handed helix presenting a deep minor groove and no discernible major groove (2).

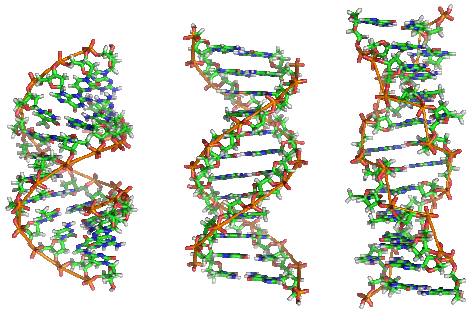


**Figure I.4:** Three-dimensional structures of A- (left), B- (middle) and Z- (right) DNA. Orange bonds highlight the spatial arrangement of phosphate groups and do not represent chemical bonds. http://www.mun.ca/biology/scarr/A_B_Z_DNA.html (May 2010)

During various life-cycles, hereditary information is replicated and transmitted *via* meiosis and mitosis. In accordance with Watson and Crick's central dogma the genetic information (genotype) of any living organism is either stored as DNA replicated by specific enzymes into a complementary strand or is transcribed by other enzymes into RNA (mRNA or the 'messenger' RNA). RNA is further processed by ribosomes that translate a succession of three nucleotides (or 'codon') into an amino acid, according to a universal genetic code. In this way, whole nucleotide sequences are translated into amino-acid sequences (peptide and proteins) that form the molecular phenotype of each individual.

## *I.2 DNA precursors in vivo*

In all organisms the *in vivo* precursors of DNA are deoxyribonucleoside triphosphates (dNTPs) and are derived *in situ* from their corresponding ribonucleoside triphosphates (NTPs). A NTP is typically composed of the corresponding ribonucleoside monophosphate linked *via* the 5′-phosphate group to a pyrophosphate moiety. Three consecutive phosphate groups are referred as α-, β- and γ-phosphates, as represented in Figure I.5.

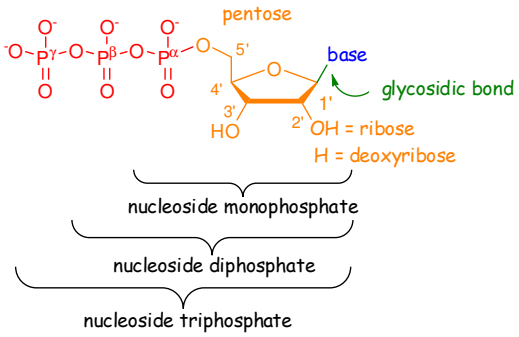



**Figure I.5:** Molecular structure of nucleoside triphosphates with conventional nomenclature of phosphate groups.

The biosynthesis of NTPs starts with the assembly of simple fragment molecules that undergo successive biochemical transformations carried out by specific enzymes (3). Such process consists in a series of discrete/individual steps that differ whether a purine or a pyrimidine species is being synthesized. *De novo* synthetic purine and pyrimidine pathways respectively yield inosine and uridine monophosphates (I- and U-MP). Base differentiation towards the monophosphate species of adenosine (AMP) or guanosine (GMP) on one side, and cytidine (CMP) or thymidine (TMP) on the other side occurs *via* specific enzymes. Once the corresponding nucleoside diphosphates are formed with the help of specific kinases, they are further reduced to deoxynucleoside diphosphates by ribonucleoside diphosphate reductase. Finally, base-specific kinases phosphorylate the diphosphates, resulting in the series of triphosphates dATP, dGTP, dCMP and dTTP (Figure I.6).

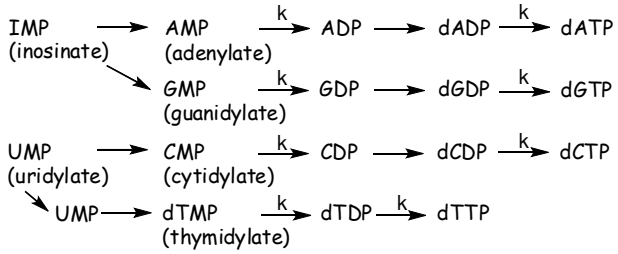



**Figure I.6:** Biosynthesis of deoxyribonucleosides triphosphates: a cascade of enzymatically catalyzed transformations yielding four DNA building blocks. Kinases are symbolized by 'k'.

DNA building blocks can also be regenerated from their successive metabolites, although not in all growing cells, according to what is called the salvage pathway (3).

## *I.3 Polymerization enzymes*

The transmission of the genetic information is crucial for both survival and evolution of all living species. Polymerases were identified as the enzymes endowed with replication activity and have been therefore the subject of intense studies (4;5).

### I.3.1 History and classification

Since the discovery of DNA polymerase I of *Escherichia coli* by Kornberg in 1955 (5), several polymerases have been identified in prokaryotes and eukaryotes. Traditionally polymerases are classified according to their substrate or template specificity. In this classification two major groups of polymerases can be distinguished: the DNA and the RNA

polymerases. Both groups can then be subdivided into a template-dependent and a template-independent class. A further distinction can be made between DNA-dependent (such as Klenow fragment and Vent polymerase) and RNA-dependent polymerases (reverse transcriptases (RT) such as HIV-1 RT and AMV-RT and eukaryotic telomerases). Similarly, RNA polymerases can either be template-dependent, such as T7 bacteriophage polymerase and or NS5B polymerase from hepatitis C virus) or template independent (e.g. poly(A) polymerase).

A second classification which is most commonly used distinguishes polymerases on the basis of their amino acid sequence homology. This allows their subdivision into seven different families (A, B, C, D, X, Y and RT, Table I.1). Family A is also referred to as the DNA polymerase I (pol I) family and includes the Klenow fragment, Taq polymerase, mitochondrial γ and polymerases from odd-numbered phages (T3, T5 and T7). It is probably the most studied polymerase family. The B family includes all eukaryotic replicating DNA polymerases, the archeon Vent polymerase, Pfu polymerases and the polymerases form the phages T4, 46 and RB69. The DNA polymerases that duplicate bacterial chromosomes are members of the C family of polymerases. These particular polymerases share no sequence homology with any of the other families (6). Family D contains most of the archeon polymerases. Family X is also called the terminal transferase family and gathers the polymerases β, λ and μ. These enzymes are not directly involved in replication but are specialized in tasks such as DNA repair or RNA maturation. Polymerases involved in mutagenesis such as the mammalian polymerases η, ι, and κ are members of the polymerase Y family. They lack 3′-5′ proofreading activity. As a consequence they replicate DNA with low fidelity *in vitro* or play a role in the by-passing of damaged bases in the replication process. Finally, the reverse transcriptase (RT) family includes the eukaryotic telomerases and the viral reverse transcriptases. Viral RTs also lack 3′-5′ proofreading activity (7-9), and replicate DNA with low fidelity, endowing corresponding viruses with high mutation rates.

| **Table I.1:** Classification of polymerases according to their amino acid sequence homology | |
|---|---|
| Family A | Klenow, Taq, γ, T3, T5 and T7 DNA polymerases |
| Family B | α, δ, ε, Vent, Pfu, T4, T6 and RB69 polymerases |
| Family C | α subunit of DNA pol III |
| Family D | Archeon polymerases |
| Family X | β, λ and μ polymerases |
| Family Y | η, ι, and κ polymerases |
| Family RT | HIV-1, HIV-2, MLV and AMV RT, telomerases |

In the process of enzymatic DNA synthesis, DNA polymerases are confronted to a pool of four structurally similar dNTPs from which they must select the correct substrate for incorporation into the DNA duplex. Every second, several hundreds of dNMP are inserted with a fidelity varying among polymerases. Error rates for the replication process vary from 1 in 1,000,000 bases for the high fidelity polymerases (eukaryotic DNA replication), to 1 in 1,000 for lower fidelity polymerases involved in viral DNA replication. Polymerases that by-pass DNA lesions go down to making 1 error in 10 insertions (4).

### I.3.2 Structure and mechanism

In spite of their fundamental role, DNA polymerases are different among diverse life forms. Nevertheless, important functional regions are highly conserved among different classes (10) and the overall shape of all polymerases is similar. Their three-dimensional structure resembles a right-hand, consisting in 'fingers', 'palm' and 'thumb' domains (Figure I.7). Although significant sequence variations are observed among the fingers and thumb domains of different polymerases, the respective functions of these domains remain universal. The fingers are crucial for nucleotide accommodation, insertion fidelity, and for dsDNA-enzyme complex stability. The thumb is responsible for the positioning of the replicating duplex, the processivity of the synthesis and plays an important role in the translocation process. The residues of this domain form extensive interactions with the phosphate backbone along the minor groove of the DNA duplex. The palm domain is the most conserved region and contains the active site responsible for the chemical step of nucleotide incorporation into DNA. A catalytic pocket accommodates the sugar moiety and allows steric discrimination

between deoxyribo- and ribo-nucleoside. Three conserved carboxylate residues play a crucial coordinating role towards the triphosphate moiety.

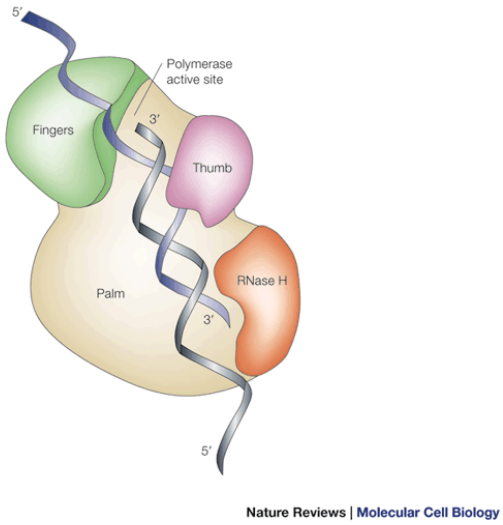



**Figure I.7:** Schematic representation of the 'palm', 'fingers' and 'thumb' domains of HIV-1 reverse transcriptase (reproduced from (11)).

Polymerases catalyse DNA synthesis in a template-dependent manner. A DNA duplex made of a short primer DNA hydridized to a longer DNA template is recognized by the polymerase. At the active site, upon accommodation of a dNTP, the DNA polymerase promotes the nucleophilic attack of the 3′-hydroxyl group of the primer on the α-phosphate of the incoming dNTP. This reaction called the 'phosphoryl transfer reaction' is thermodynamically driven by the dissolution of the phosphate bond, the formation of a more stable phosphodiester bond, and the subsequent hydrolysis of the pyrophosphate moiety. In the transition state, the α-phosphorous is thought to be pentacoordinated (Figure I.8) (12). Furthermore, two magnesium ions (A and B) are thought to be tightly bound to three carboxylate functions belonging to the active site of the enzyme. Two of those residues consist of aspartic acids and are widely conserved in the polymerases. Magnesium ions A and B are also interacting with all three phosphate groups. The catalytic ion A is thought to activate the 3′-hydroxyl group for the attack, acting as a Lewis base. The second catalytic ion complexes the β and γ phosphate groups and is therefore thought to neutralize the negative charges of the dNTP in the initial state and during the transition state. In the final state, the released pyrophosphate moiety is likely to be leaving the active site together with catalytic metal B. Hydrolysis of the released pyrophosphate, by inorganic pyrophosphatase recruited by eukaryotic and viral polymerases (3) or by the pyrophosphatase domain of archeon

polymerases (13), is thought to provide an additional driving force for polymerization and makes the reaction irreversible.

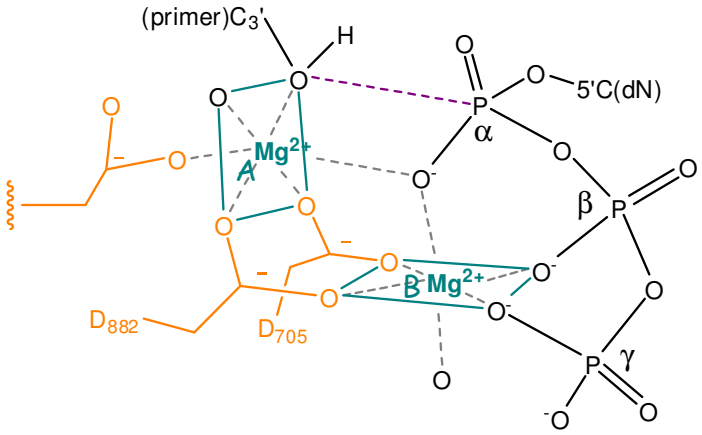



**Figure I.8:** The two-metal ion mechanism of the phosphoryl-transfer reaction catalyzed by *E. Coli* DNA polymerase I (adapted from (12)). The carboxylate residues are shown in orange the coordination bonds of the magnesium ions in grey, and the bond in formation in purple.

Extensive kinetic studies and numerous structural studies of polymerase complexes have helped in understanding the accuracy of DNA replication. Mechanistic and structural features that are common to almost all polymerases allow to control the fidelity of nucleotide insertion during five discrete steps. At first the dsDNA binds to the polymerase. Immediately after, the dNTP binds to the dsDNA:enzyme complex forming a ternary complex. A conformational change follows from an 'open' conformation to a 'closed' conformation (14). Finally the chemical step of phosphodiester bond formation occurs and results in the release of pyrophosphate. A subsequent conformational change allows opening of the active site and translocation of the enzyme to the new 3′- hydroxyl primer terminus.

Five interaction sites of importance were identified during the course of the nucleotide insertion process. The template base starts its journey at a 'pre insertion site' of the polymerase. During a conformational change that allows the enzyme to go into a 'closed' conformation, the template base is positioned in the 'insertion site'. As a consequence a hydrophobic pocket forms around the deoxyribose and the base of the dNTP and the negatively charged phosphate residues are positioned between hydrophilic residues of the enzyme. At the 'insertion site' the dNTP base forms hydrogen bonds with the opposite template base. In the open conformation the template base is occluded by the aromatic side chain of a residue of the enzyme, often a tyrosine, that stacks with the template base. In the subsequent 'closed' conformation the residue is displaced, giving place to the dNTP base and

allowing base pairing. Following formation of the new phosphodiester bond, the newly formed base pair moves to the 'post insertion site' while the pyrophosphate leaving group is released and the enzyme changes back from the 'closed' to the 'open' conformation. In the case of a processive enzyme, the ternary complex remains in place and the following nucleotide insertion proceeds. Otherwise, the polymerase is said to be distributive and the ternary complex can be disrupted and a new ternary complex has to form, resulting in substantial slowing down of the DNA synthesis (5;15;16).

### 1.3.3 HIV-1 reverse transcriptase

Due to the fact that the viral enzyme HIV-1 reverse transcriptase is a drugable target for anti HIV-1 therapies, it has been the subject of many investigations. Thanks to the availability of crystal structures in various liganded states, a lot of information is available about structure-activity relationships. HIV-1 RT is a heterodimeric enzyme composed of a 66 kDa unit (p66) endowed with polymerase activity and of a 51 kDa unit (p51) possessing RNase H activity (Figure I.9). HIV-1 RT belongs to the RNA-dependant DNA polymerases. This polymerase features the typical right-hand overall 3D structure (Figure I.7) and, as for all its congeners, the active site for nucleotide incorporation is situated in the palm domain. Both the fingers and the RNase H domain are responsible for primer:template (P:T) positioning. The three carboxylate functions essential for catalytic activity are situated at residues Asp110, Asp185 and Asp186. HIV-1 RT preferentially functions in the presence of magnesium ions (17).

HIV-1 RT (referred to as RT as from this section of the text) lacks a proof-reading function and is therefore known to be error-prone, making one misincorporation every 1000 incorporations. As a consequence, it has a high mutation rate (18). It also has broad substrate specificity and is therefore able to incorporate many dNTP analogues into DNA. Current highly active antiretroviral therapies (HAART) include at least two compounds that inhibit the polymerase function of RT and thus interfere with the synthesis of viral DNA. The three-dimensional structure of RT bound to dsDNA and diverse 3'-deoxynucleotides was elucidated by X-ray crystallography (17). The tremendous amount of data gathered since its first characterization and the low substrate specificity make it a very good tool to study the mechanism of dNTP incorporation and to test new dNTP analogue candidates as substrates.

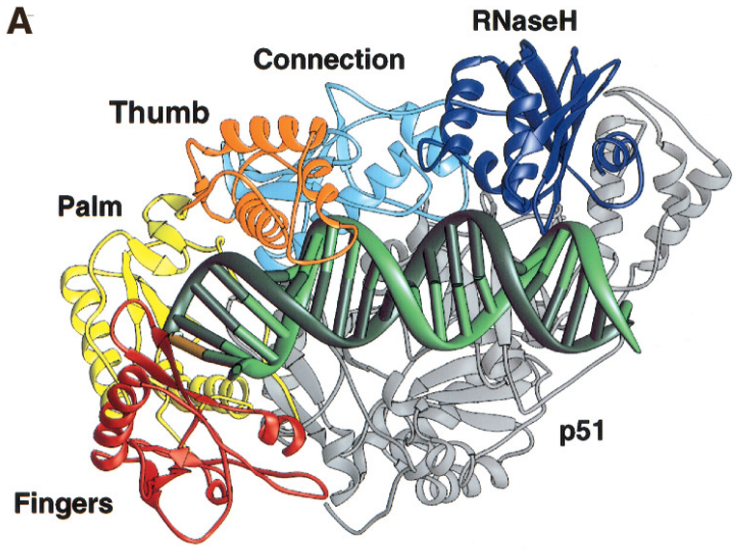



**Figure I.9:** The structure of the HIV-1 RT catalytic complex with the polymerase active site on the left and the RNase H domain on the right. The domains of p66 are in colour: fingers (red), palm (yellow), thumb (orange), connection (cyan), and RNase H (blue); p51 is in gray. In the two chains, the domains have very different relative orientations. The DNA template strand (light green) contains 25 nucleotides, and the primer strand (dark green), 21 nucleotides. The dNTP is in gold. This picture was reproduced from (17).

## *I.4 In vitro chemical synthesis of nucleic acids*

### I.4.1 Oligonucleotide synthesis

With the aim to studying DNA and RNA secondary structures and conformations and their biophysical properties, chemists have developed synthetic procedures for the small scale production of oligonucleotides of variable lengths. In the 1960s, oligonucleotide (ON) synthesis was carried out in solution. Chain elongation resulted from block-synthesis following the phosphotriester method. This method suffered a lack of reproducibility and sequence precision; in addition, the process was strenuous, implicating silica gel column chromatography purification after each coupling. In order to overcome those problems, solid-phase synthesis was developed (19). Nowadays the solid support consists mostly in a glass carrier functionalized with primary amines, namely controlled pore glass (CPG). A succinate group serves as linkage between CPG and the first building block. Using this method oligonucleotide synthesis is fast, reliable, reproducible and allows automation. Three synthetic strategies are available for solid-phase synthesis. The phosphite triester or phosphoramidite method adapted to solid support by Matteucci and Caruthers is the most

employed (20) and is represented in Figure I.10. The key step of this water sensitive $P^{III}$ chemistry is a coupling between a tetrazole activated dN phosphoramidite and the 3′-hydroxyl group of the starting oligonucleotide which offers the major advantage to be very high yielded. Oxidation is usually carried out with the use of iodine in the presence of aqueous pyridine. Alternative methods are available for non aqueous and/or non-alkaline oxidation (21;22).

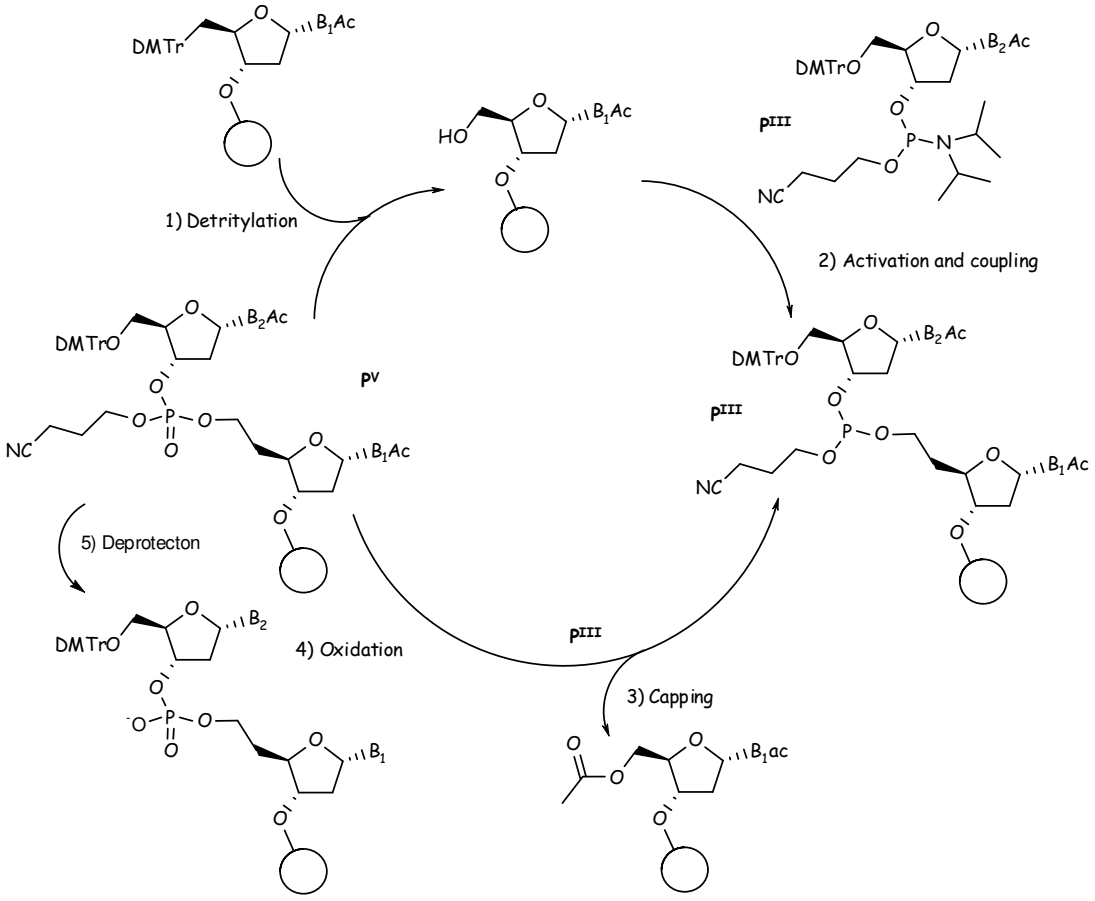



**Figure I.10:** Schematic representation of the synthesis cycle for oligonucleotide synthesis following the phosphite triester method: 1) The 3′-end nucleotide carrying base $B_1$ is linked to the CPG *via* a succinyl linker (both represented by a disk) and protected by a dimethoxytrityl (DMTr) group, 2) the phosphite triester carrying base $B_2$ is activated and coupled to the pre-existing oligonucleotide, 3) unreacted oligonucleotide are capped *via* acetylation, 4) the phosphorous in oxidation state +III is oxidized towards oxidation state +V, yielding a phosphotriester, 5) the full length ON is deprotected and cleaved from CPG by strong basic treatment.

The H-phosphonate method, represented on figure I.11, also involves $P^{III}$ chemistry. It proceeds *via* pivaloyl chloride activation of the building block's phosphate function and offers a shorter coupling time. Since the oxidation step is carried out after completion of strand synthesis, it is possible to introduce phosphate modifications during this step (for example, sulphur oxidation yields a fully phosphorothioated ON).

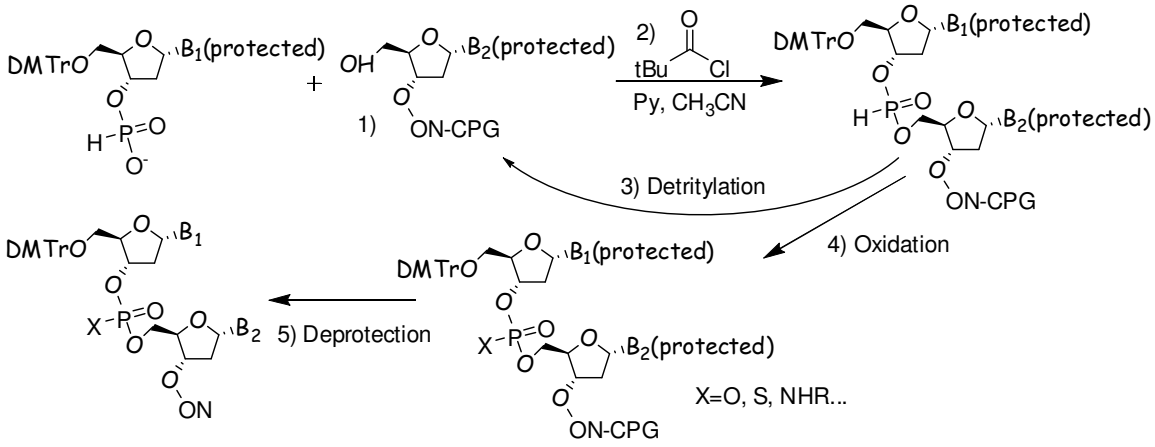



**Figure I.11:** Schematic representation of the synthesis cycle for oligonucleotide (ON) construction following the H-phosphonate method: 1) The 5′-end nucleotide carrying base $B_2$ is linked to CPG, 2) the H-phosphonate carrying base $B_1$ (protected by a DMTr in 5′ position) is activated and coupled to the pre-existing oligonucleotide, 3) the 5′ end of the ON is deprotected *via* acidic treatment and re-enters the synthesis cycle, or 4) the backbone phosphorous in oxidation state +III is oxidized by iodine towards oxidation state +V, yielding a phosphodiester or a phosphate modified backbone 5) the full length ON is deprotected *via* basic treatment.

Finally, the phosphotriester method uses a sulfonyl chloride species in combination with methylimidazole as activating agent of a protected 3′-phosphorylated dN building block, allowing subsequent attack of the phosphate by the 5′-hydroxyl group of the pre-existing oligonucleotide. Regardless of the synthetic method used, the resulting oligonucleotide is purified *via* a combination of preparative reverse phase (RP) and/or ion exchange (IE) chromatography.

### I.4.1 Synthesis of DNA building blocks

Methods which rely on a similar chemistry to the one presented in the previous paragraph can also be used to synthesize nucleotide triphosphates (dNTPs) and their analogues (an overview is given in Figure I.12). Among those, Ludwig and Yoshikawa's one pot synthesis is the most

straightforward method. A deoxynucleoside is first phosphitylated by phosphorous oxychloride (23). Subsequent coupling to pyrophosphate tributylammonium yields a dNTP (24). Another powerful method consists in activating the phosphate group of a dNMP. Smith and Korana reported the synthesis of dNTP from dNMP and pyrophosphoric acid using dicyclohexylcarbodiimide (DCC) as coupling agent (25). Abraham *et al.* also used DCC activation to couple dNMPs to amino acids (26). DCC activation allows the high yielded synthesis of more stable activated species like (2-methyl) imidazole-dAMP (27) or morpholine-dAMP (28) and subsequent coupling *via* nucleophilic substitution by dNMPs and alcohols (29;30). Deoxynucleotide imidazolates can be stored and are therefore very convenient to use.

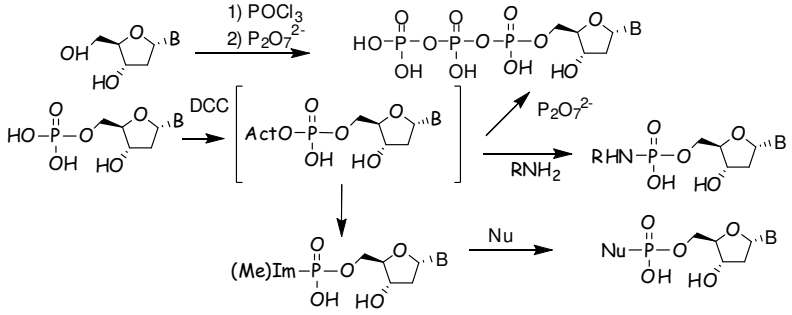



**Figure I. 12:** Overview of principal synthetic routes towards dNTPs and their analogues. 'Act' symbolizes an activation group.

The Mukayiama oxidation-reduction condensation (Figure I.13) offers an efficient alternative (31). In this method dipyridinium disulfide (DPDS) is used as a condensing agent and triphenylphosphine ($PPh_3$) serves as reducing agent.

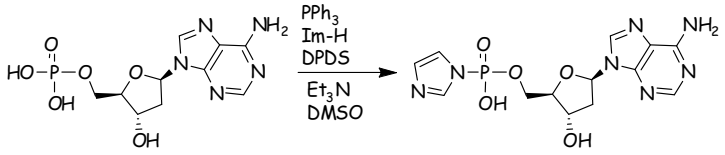



**Figure I.13:** Synthetic scheme of the Mukayiama oxidation-reduction condensation.

Alternatively, the H-phosphonate approach consists in forming the H-phosphonate monoester of a deoxynucleoside *via* transesterification of diaryl H-phosphonate (e.g. diphenyl H-phosphonate (32)) with subsequent hydrolysis of the second aryl group, or directly *via* reaction with phosphonic acid (33). Amines can be phosphorylated by diaryl or dialkyl phosphonates in the presence of carbon tetrachloride as reported by Atherton *et al.* (34).

Subsequent addition of hydroxylic compound to the H-phosphonate monoester, followed by activation with pivaloyl chloride (35), yields a H-phosphonate that can be oxidized with iodine (36) or other strong oxidants (37;38). An overview of these methods is given in Figure I.14.

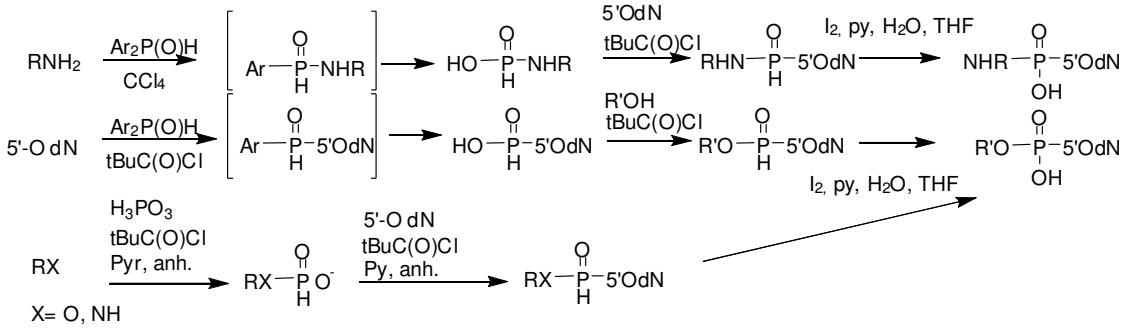



**Figure I.14:** Overview of principal H-phosphonate synthetic routes towards dNTP analogues.

Finally, in the phosphoramidite or phosphite triester method, phosphorous trichloride is derived in multiple steps into a deoxynucleoside phosphoramidate (Figure I.15). Subsequent activation by a large choice of activating agents (reviewed in (39)), followed by coupling to diverse nucleophilic species (using tetrazoles or triazoles as coupling agents) and final oxidation in appropriate conditions yields the desired dNTP analogues. This method involves extremely reactive chemicals and necessitates anhydrous synthetic conditions. It is therefore the least bench-friendly.

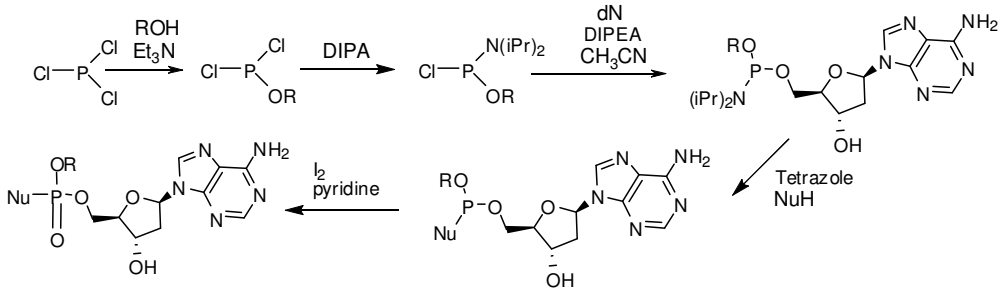



**Figure I.15:** General scheme for synthesis of dNTP analogues *via* phosphite triester method.

## *I.5 Applications of nucleic acids*

Oligonucleotides are widely investigated for their use as pharmaceuticals. Diverse backbone modified oligonucleotides have been developed for the purpose of a therapeutic method later called the antisense approach (40), where a defined sequence binds to a pathogenic mRNA

and in this way inactivates it. The most realistic antisense therapeutic applications of oligonucleotides for antiviral purposes were reviewed by Van Aerschot in 2006 (41). The first generation of antisense oligonucleotides (AS-ONs) was modified at the phosphate position exclusively and the most successful were found to be phosphorothioate oligonucleotides (PS-ONs) (Figure I.16).

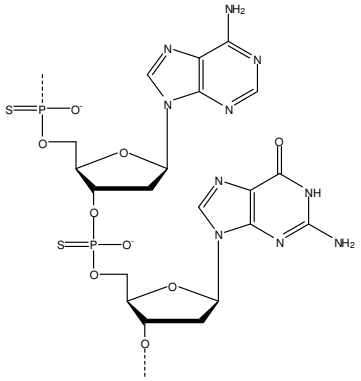


**Figure I.16:** Fully phosphorothioated oligonucleotide.

The second and third generations of AS-ONs consist in sugar modified oligonucleotides (e.g. peptide nucleic acid, or PNA (42), cyclohexenyl nucleic acid, or CeNA (43), phosphorodiamidate morpholino oligonucleotides or PMOs (Figure I.17).

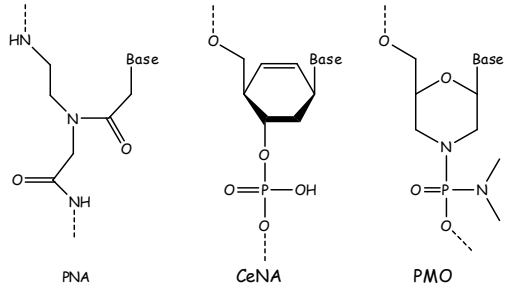



**Figure I.17:** Molecular structure of the backbone of peptide, cyclohexenyl and phosphorodiamidate morpholino nucleic acids.

Other therapeutic applications include RNA interference, aptamers technology (44) or immunostimulatory strategies (45). Nucleic acids have also found many applications in biochemical research. They are used as 'primers' and 'probes' in the polymerase chain reaction (PCR) technology (46) (involving biotinylated and fluorescent probes), DNA sequencing (47;48), and high-throughput screening (49;50) (using oligonucleotides in microarrays). In addition, site specific mutagenesis is used to study structure-activity relationships of proteins. Finally, nucleic acids may be used in nanotechnologies as

components for the construction of nanostructured systems by virtue of the structural properties and conformational dynamics of the self-assembling double helix (51;52), or of the electrostatic status of the backbone (53). An example of many possible geometric structures which could be generated is shown in Figure I.18.

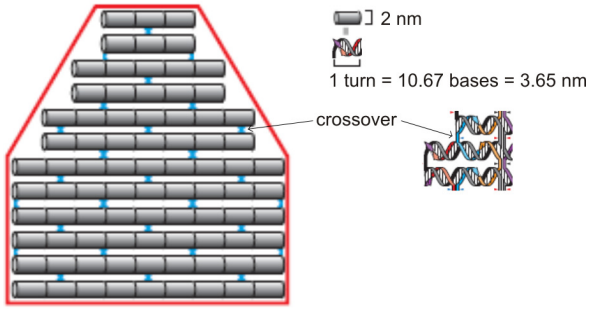



**Figure I.18:** A nanoscale shape (in red) composed of parallel double helices joined by periodic crossovers (in blue) (reproduced from (52)).

The advances that have followed the control of nucleic acid synthesis have led to the proliferation of genetically modified organisms (GMOs) for research and industrial purposes, which possess synthetic nucleic acid constructs in their genome, endowing them with unusual properties or functions. Ten years ago, a lot of hope was given to the developing world at the presentation of the so-called 'golden rice' which had been engineered with additional nutritional properties in order to fight malnutrition issues (54). Until now, only intensive agriculture has benefited of engineered crops with enhanced resistance to viruses, fungus or drought. The agro-industrial giant Monsanto has been successful in manipulating corn genome and growing it at an industrial scale. There is good hope that some plant species could be modified so as to produce drugs or vaccines (55). However, since the genetic alphabet and genetic code are uniform and universal, it has become a major topic of debate that one cannot avert pollution of wild ecosystems by those genetically modified species. In this regard, GMOs might be threatening biodiversity on the long run. In fact, 'polluted' ecosystems were detected around GMOs fields, showing the lack of control of engineered nucleic acid dissemination. So far, experts have been unable to develop GMOs that do not disseminate their unusual genetic material in their environment. Therefore, there is an urge to develop safe genetically modified organisms whose hereditary information cannot interfere with the one of wild species. It could be envisaged that an independent molecular factory is introduced in natural organisms and participates in this manner in resistance mechanisms or in the production of exogenous molecules. In this regard, the notion of 'xenonucleic acids'

(XNA) that differ, in both their base alphabet and their backbone structure, from natural nucleic acids (DNA and RNA), has been introduced recently by Herdewijn and Marlière as part of an 'orthogonal episome' (56).

## *I.6 Orthogonal DNA synthesis in vivo*

### I.6.1 Presentation of the concept

The ubiquity of (deoxy)nucleoside triphosphates as nucleic acid precursors and of DNA/RNA polymerases as polymerization enzymes implies that, if a precursor analogue is substrate for one given polymerase, stakes are high that it will also function as substrate for foreign polymerases (different cell type, different organism) with various efficiencies. For example it is a major concern that nucleoside-like RT inhibitors (NRTIs) also disrupt mitochondrial DNA synthesis and that their therapeutic effect is concomitant with many undesirable side effects (57;58). In addition, artificial DNA synthesis and DNA manipulation as technologies for engineering of genetically modified organisms often raise concerns of interference with natural systems. The term 'orthogonality' suggests that two systems are mutually exclusive. Thus, 'orthogonal' DNA synthesis *in vivo* aims at segregating the artificial synthesis from that of natural organisms. In the frame of a synthetic biology project of genetic enclave (56), as shown in Figure I.19, the use of alternative building blocks (modified sugar, base and leaving group) and of a specific polymerase (mutated and selected for its specificity towards the new building blocks) could allow the segregation from the natural deoxynucleoside triphosphates and avoid interference with natural genetic systems. In order to modify one single parameter at a time, all three chemical approaches (sugar, base or leaving group modification) need to be investigated separately. Polymerases specific for each modified unit may be mutated towards recognition of each modified unit separately. Directed evolution could then culminate with the final selection of a polymerase specific for the fully modified nucleic acid precursors.

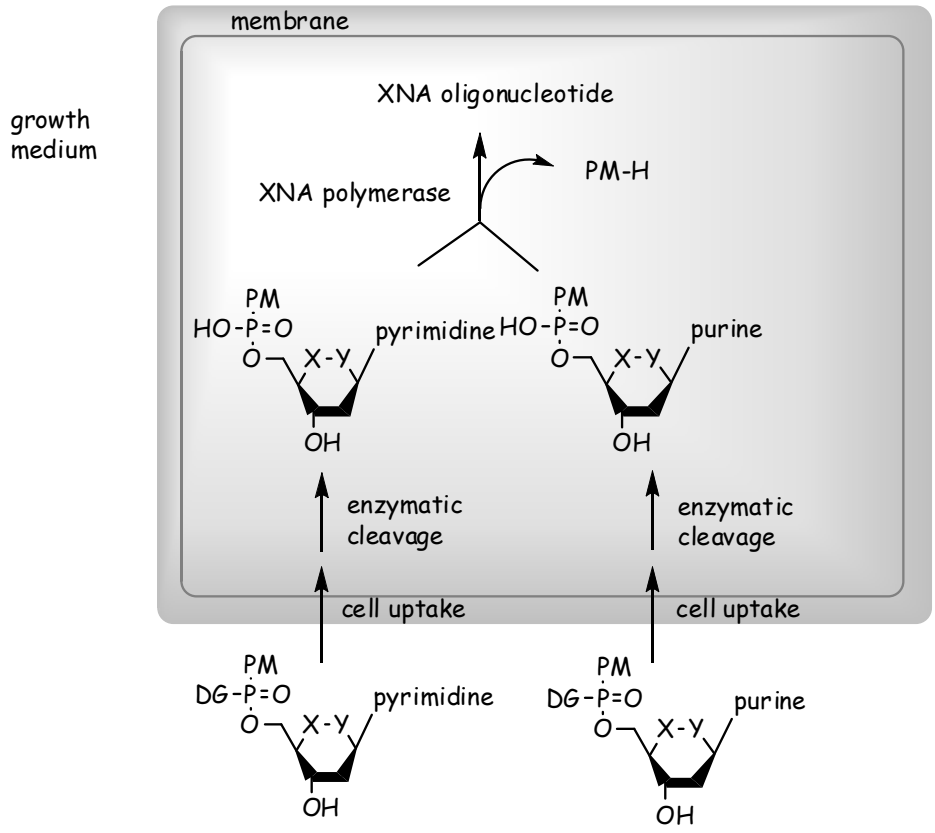



**Figure I.19:** Schematic representation of orthogonal DNA synthesis based on 5′,6′-modified hexose building block (adapted from (56)). PM= pyrophosphate mimic, DG= delivery group.

## I.6.2 Monomer diversification

During the last ten years the search for an alternative phosphate-sugar backbone offering identical stability and replication machinery has led to the discovery of diverse dNTP analogues containing a modified sugar moiety. Among others, cyclohexenyl nucleic acids analogues (CedNTP) (59) and hexitol nucleic acids (HNA) (60) are possible alternatives. In their molecular structure the (2′-deoxy)ribose sugar unit of the backbone is replaced by a cyclohexenyl moiety or an hexitol sugar, respectively. CeNA and HNA are able to form stable single- and double-stranded structures with DNA. The corresponding triphosphates (Figure I.20) are processed by usual DNA enzymes. They are, therefore, not orthogonal on their own.

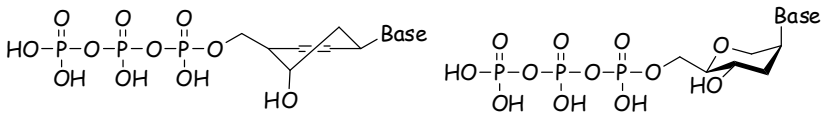



**Figure I.20:** Cyclohexenyl nucleoside triphosphate, hexitol nucleoside triphosphate.

Expanding the genetic alphabet by creating novel base pairs is becoming a great challenge since various properties like specific base pairing, stacking, specificity, and selectivity are required for replication fidelity. In addition, extra bases should not be able to form base pairs with the natural ones. In this context, the Romersberg group introduced several heteropairs of bases, whose pairing is based on complementary hydrophobicity (61) and could successfully employ some of them (Figure I.21) for PCR (62) and transcription (63).

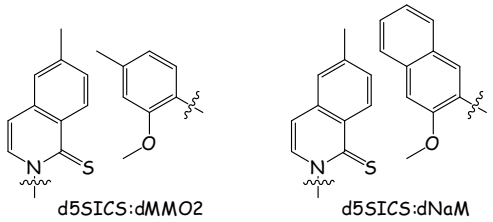



**Figure I.21**: Molecular structure and base pairing between the 1-thioisocarbostiryl moiety as a base substitute and methylmethoxyphenyl (left) or 2-methoxynaphtalene (right) as its complementary base.

Remarkably, Neumann *et al.* recently extended the orthogonalization concept to the translation step by 'expanding the genetic code'. An engineered orthogonal ribosome was able to encode 4-base codons into unnatural amino acids (64).

### I.6.3 Enzyme evolution

The issue of orthogonality, in particular enzyme specificity and fidelity, can be addressed simultaneously by developing appropriate DNA enzymes. Several *in vitro* directed evolution methods such as DNA shuffling/recombination (65) using affinity selection techniques such as phage-display (66) and compartmentalized self-replication (CSR (67), depicted in Figure I.22), are now available to direct the evolution of a given enzyme towards enhanced substrate specificity (68). The generated mutant will therefore segregate natural DNA precursors and process non-natural nucleic acids ('xenonucleic acids', XNA) precursors in exclusivity. Subsequent use of bacteria like *Escherichia Coli* as host for an orthogonal machinery dedicated to synthesis of informational polymers could provide a direct and unlimited access to tailor-made XNA for diverse therapeutic or analytic applications.

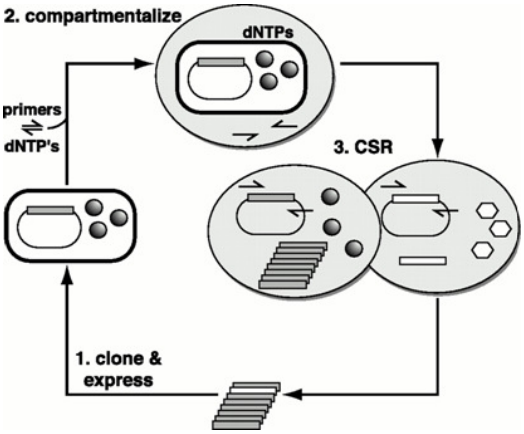



**Figure I.22:** Schematic representation of compartmentalized self-replication. 1) A repertoire of polymerase genes is cloned and expressed in *E. coli*. Spheres represent active polymerase molecules. 2) Bacterial cells (white rectangular unit) containing the polymerase and encoding genes are suspended in reaction buffer containing flanking primers and dNTPs and segregated into aqueous compartments (in grey). 3) The polymerase enzyme and encoding gene are released from the cell, allowing self-replication to proceed. Poorly active polymerases (white hexagons) fail to replicate their encoding gene. The offspring polymerase genes of active polymerases are released and cloned for another cycle of CSR (reproduced from (69)).

### I.6.4 Modification of the leaving group

An unprecedented approach in the search for orthogonality consists in replacing the chemical moiety of nucleic acid building blocks which serves as leaving group in the polymerization mechanism by a mimicking group. The leaving group of natural DNA precursors is a pyrophosphate (PPi) and consists in the β- and γ-phosphate moiety of the nucleoside triphosphate (the universal denomination of these phosphate groups is illustrated in Figures I.5 and I.8). Pyrophosphate serves as universal biochemical energy storage and is therefore ubiquitous in living organisms. The tremendous amount of data available on DNA polymerases and their polymerization mechanism enabled to characterize the major characteristics and the structure-function relationship of PPi in this process (12;17), setting some of the requirements for rational design of potentially successful pyrophosphate mimic candidates endowed with similar efficiency. For example, in the laboratory of Goodmann modifications have been introduced on the bridging oxygen between the phosphorous atoms in α and β positions. Such modifications affected the polymerization rate ($k_{pol}$) and replication

fidelity. Interestingly, variations in $k_{pol}$ could be correlated to the corresponding variations in the pKa of the departing biphosphonic acid (70;71).

In order to function as leaving group during the polymerization process, a candidate should be able to behave chemically and biochemically in a similar way to pyrophosphate. Molecular models, based on crystal structures of various liganded states of polymerase, suggest that a leaving group should fit into the active site pocket, therefore limiting the steric size of the leaving group. They also imply that the leaving group should be able to form hydrogen bonding (triphosphate and sugar moiety) and Van der Waals interactions (base and sugar moiety) with the enzyme's residues and with the catalytic ions. From a strict chemical point of view, the leaving group's structure should promote the nucleophilic attack on the α-phosphate (Figure I.8). Therefore it should be linked to the phosphate *via* a highly polarisable bond. The released molecule should also be basic enough to be readily protonated after departure.

A first possibility is to replace the linking oxygen atom with various heteroatoms that would generate good departing anions upon cleavage from the α-phosphorous. Some bioisosteres of oxygen that could fulfil these requirements are nitrogen and sulphur (72). Another possibility is to conserve the linking oxygen atom and substitute it with various chemical groups whose corresponding alcoholates should be good departing anions. As previously presented on Figure I.8, the coordination of two magnesium ions is thought to play an important role in the consensual nucleotide incorporation mechanism. As a consequence, it seems to be of utmost importance that the leaving group candidate possesses good coordinating capacity. Carboxylic acid, sulfonic acid, sulphonamide and aminophosphate are commonly used in drug design as the less bulky bioisosteres of the phosphate functional group. The rationale for the design of pyrophosphate mimics is summarized in Table I.2.

**Table I.2:** Rational design of pyrophosphate mimics.

| | **Lead compound** | **Bioisosteres** |
|---|---|---|
| Linkage | PPi-O-P | RO-P, RNH-P, $R_2N$-P, R-S-P |
| Electronic status | $-PO_3^{2-}$ | $-COO^-$, $-SO_3^{2-}$, $SO_2-NH^-$, $-PO_2-NH^-$ |
| Coordination bond donor | Non-bridging oxygen of $PO_3^{2-}$ | All of the above<br>Small O or N containing cycles |

Some pyrophosphoric acid analogues like phosphonoformic acid (or foscarnet (Figure I.23)) (73) or other (reviewed in (74)) endowed with antiviral activity against viruses (e.g. herpes simplex virus) are thought to mimic pyrophosphate by coordinating specific divalent ions essentials for the functioning of many viral enzymes, exampling the possibility of other chemical functions to mimic pyrophosphate in this regard.

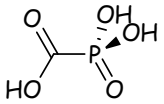



**Figure I.23:** Molecular structure of phosphonoformic acid (foscarnet).

The following issue in the replacement of pyrophosphate by an alternative leaving group is the fate of the released molecule in the biological system. Similarly to PPi, in order to promote advancement of the polymerization reaction and to prevent accumulation in the environment, an alternative leaving group should be preferably metabolized into simple fragments and/or possibly recycled for further biosynthesis. Importantly, the PPi mimic and corresponding metabolized fragments should bear little or no toxicity for the cellular/bacterial environment. If a chemical group meeting all previous requirements is developed, it could be possible to use it in order to orthogonalize the substrate recognition process. This in turn allows the segregation of natural deoxynucleoside triphosphates and avoids interference with natural genetic systems.

It is essential to mention that such candidate leaving groups could be valuable in the search for new antiretroviral compounds. So far, most nucleoside-like polymerase inhibitors developed for anticancer or antiviral therapies are delivered *in situ* in their nucleoside or nucleoside monophosphate form and require further phosphorylation operated by cellular kinases (75-77). Cellular kinases unfortunately have greater substrate specificity and do not process some nucleoside analogues. For example, dideoxyuridine (ddU) is a very poor substrate for the host's thymidine kinases (78). Yet, once triphosphorylated, this dideoxynucleoside is a potent antiviral compound as its administration leads to termination of viral DNA synthesis. Prodrug approaches involving the substitution of dNMP analogues with a chemical masking group have been developed in order to by-pass the first phosphorylation step and simultaneously allow the monophosphate species to enter the cell (79). Surprisingly, the mimicking of pyrophosphate by a structurally similar group has been little investigated for the purpose of by-passing the kinase pathway. The administration of a ddNMP analogue

substituted with a pyrophosphate mimic as masking group (and if necessary a lipophilic masking group for cellular uptake (80)), could allow the by-passing of the phosphorylation cascade (77), as shown in Figure I.24.

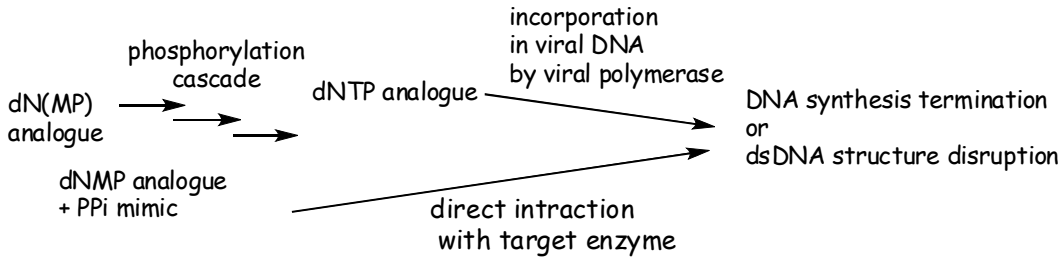



**Figure I.24:** By-pass of phosphorylation cascade of deoxynucleotide analogues with the use of a deoxynucleotide monophosphate substituted with a pyrophosphate mimic.

## *I.7 Aims of the study*

In enzymology the investigation of modified substrate recognition by enzymes is a great tool to gain valuable information about structure-activity relationships. In particular, the influence of the modification of a defined part of the substrate molecule on the polymerase function allows gaining insight on the polymerization mechanism. Efficient substrate mimics can be used in the development of antiviral strategies. On the other hand, modified oligonucleotides have numerous potential applications in therapeutics and diagnostics. Unfortunately, their chemical synthesis is laborious and costly, and the length of the products is limited. Therefore the pharmaceutical industry could benefit of alternative ways to produce modified oligonucleotides, such as the potential enzymatic synthesis presented in our study.

The present study focuses on the modification of the pyrophosphate moiety of natural deoxynucleoside triphosphates and its influence on DNA polymerization efficiency. In **chapter two**, various structural candidates, which represent potential mimics for PPi in the polymerization reaction, are synthetically coupled to natural deoxynucleotides. The capacity of the resulting dNTP mimics to act as substrates for a polymerase is investigated. The influence of the coordinating potential of the PPi mimic on the polymerization outcome is explored and insights on the influence of the phosphodiester or phosphoramidate linkage between dNMP and PPi mimic are gained. The **third chapter** presents the comparison of an amino acid dAMP phosphoramidate and its phosphodiester congener as dATP mimics. The influence of the PPi mimic's structure on dAMP single and multiple incorporation efficiency

into DNA by HIV-1 RT is investigated. In **chapter four** the enzymatic incorporation of deoxyadenosine monophosphate into DNA is described when pyrophosphate is replaced by iminodiacetic acid. Modelling studies show the fitting of the modified leaving group in the active site of HIV-1 reverse transcriptase and the issue of DNA synthesis pausing and termination is discussed. The **fifth chapter** is dedicated to the development of an iminodiacetic acid-specific biosynthesis of alternative DNA precursors. Therefore, the question whether iminodiacetic acid can also mimic pyrophosphate during enzymatic phosphorylation will be discussed.

# II. Chapter two: Phosphodiester substrates for incorporation of deoxynucleotides in DNA using HIV-1 reverse transcriptase

The results reported in this chapter were published in *ChemBioChem* in 2009 under the title 'Phosphodiester substrates for incorporation of deoxynucleotides in DNA using HIV-1 reverse transcriptase'. Giraut, A., Dyubankova, N., Xiaoping, S., Herdewijn, P., (2009) *ChemBioChem* **10**, 2246-2252.

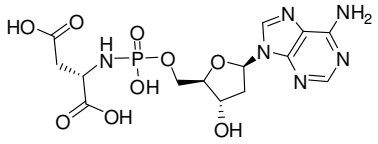
## *II.1 Introduction*

The enzymatic synthesis of nucleic acids makes use of nucleoside triphosphates as substrates and this process is driven by the release and hydrolysis of pyrophosphate. The search for mimics of the pyrophosphate leaving group, and thus for mimics of nucleoside triphosphates, was initially directed towards the discovery of polymerase inhibitors. The emergence of synthetic biology as a new research field has intensified the search for new substrates with alternative leaving groups for the polymerase-catalyzed DNA synthesis. Two important reasons for this are the dissection of nucleic acid synthesis from the energetic/metabolic/signaling systems of a cell (which also makes use of nucleoside triphosphates) and to by-pass the kinase pathway (in case sugar or base modified nucleotides are used for nucleic acid synthesis). Here we have studied different potential new leaving groups for their ability to support HIV-1 RT polymerase catalyzed DNA synthesis.

The pyrophosphate moiety has bridging (P-O-P) and non-bridging (P=O and P-OH) oxygen atoms. As it was previously assumed that the non-bridging oxygen atoms are important for the binding of the nucleoside triphosphate to the enzyme (directly and *via* metal coordination) and for efficient enzymatic polymerization of DNA, most research was concentrated on replacing the bridging oxygen with other atoms or groups.

Recently, it was demonstrated that DNA polymerases such as HIV-1 reverse transcriptase, Taq DNA polymerase and Vent (exo-) DNA polymerase, are able to process selected substrates like L-amino acid dAMP phosphoramidates (81;82). In this case the pyrophosphate

leaving group is replaced by a completely different molecule with potential coordinating properties. In particular, the L-aspartic acid phosphoramidates of deoxyadenosine (L-Asp-dAMP, Formula II.1) deoxyguanosine, deoxycytidine and thymidine monophosphates (83) demonstrated incorporation and elongation properties. The L-histidine phosphoramidate of deoxyadenosine monophosphate (L-His-dAMP) was also processed as a substrate. While L-His-dAMP is a less good substrate for HIV-1 RT than the L-Asp congener, chain elongation seems to work slightly better with L-His-dAMP. The most remarkable observation is that this error-prone polymerase is able to catalyze P-N bond cleavage. It was also demonstrated that the phosphate groups of the pyrophosphate moiety can be substituted with carboxylic acids and imidazole moieties. Furthermore it was proven that, using the L-Asp leaving group, selectivity was conserved among the four nucleobases (83).

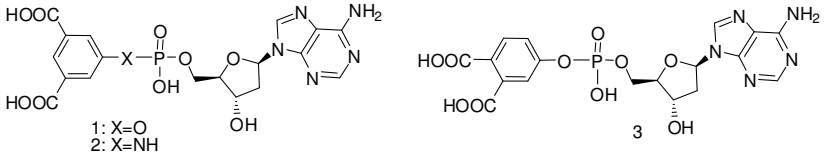



**Formula II.1**. Chemical structure of L-aspartic acid-dAMP.

Although these new building blocks are not able to compete with the natural substrates, as their substrate efficiency is lowered by a factor of $10^3$ when compared to dATP, they proved to possess the necessary structural features, size, binding properties, flexibility, and leaving group capacity to function in enzymatic DNA synthesis.

Based on the X-ray structure of a ternary complex between enzyme, template-primer duplex and substrate, a model was postulated which could explain the mechanism by which the incoming dNTP is activated by the polymerase for incorporation in the growing DNA chain (12). This model remains consensual up till now and suggests notably that one of the metal ions binds to the β- and γ-phosphates and facilitates the leaving of the pyrophosphate moiety. Garforth *et al.* could show that the interaction between a lysine residue of HIV-1 RT (Lys65) and one of the non-bridging oxygen atoms of the γ-phosphate is not necessary for the enzymatic reaction to be performed, enabling enzyme processing when only two potential chelating moieties are present (84). In this context, good ability to coordinate a magnesium cation appears to be a major requirement for a leaving group.

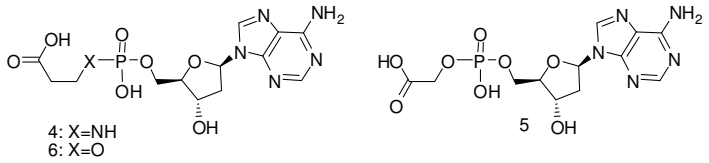
The main challenge of this project is to select a leaving group which possesses a good balance between satisfying leaving group properties, synthetic accessibility and sufficient chemical stability. Here we tested a series of new potential leaving groups in order to further explore this structure-substrate specificity. Examples of potential leaving groups with an aromatic character are given in Formula II.2. In isophtalic- (*i*PA, **1**, **2**) and phtalic- (**3**) dAMP the two carboxylic acids on the aromatic ring increase the polarity of the bond between leaving group and nucleotide and introduce the possibility to coordinate metal ions (moieties).




**Formula II.2**. Molecular structure of 5′-(isophtalic acid)-dAMP **1** and **2**, 5′-(phtalic acid)-dAMP **3**.

The influence of the replacement of the P-N linkage by a P-O linkage can be evaluated while comparing the leaving group capacity of compounds **1** and **2,** while differences observed between compounds **1** and **3** could show the influence of the position of carboxylic acid groups.

In an attempt to further simplify the structure of the leaving group, we investigated the pyrophosphate mimicking ability of a group carrying only one carboxylic acid function, for both phosphodiester and phosphoramidate analogues (Formula 3). The glycine analogue has been evaluated previously and found to be inactive as leaving group (81). Here we evaluated the oxygen analogue (**5**) of this compound, as well as the β-alanine analogue (compound **4**) and its hydroxyl counterpart, γ-hydroxypropionic acid **6**. In addition, the study of compounds **4** and **6** will allow us to estimate the role of the γ-carboxylic acid in the mechanism.



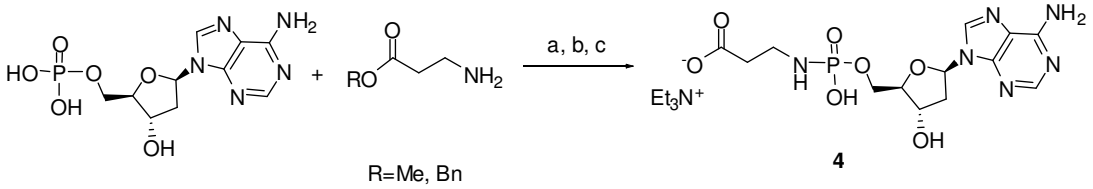

**Formula II.3.** Molecular structure of β-ala-dAMP **4**, glycolic acid-dAMP **5** and γ-hydroxypropionic acid-dAMP **6.**

## *II.2 Results*

### II.2.1 Synthesis of phosphoramidates and phosphodiesters

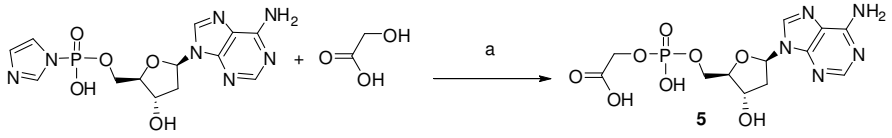
The synthesis of the methyl or benzyl esters of the protected nucleotides analogues (L-Asp-dAMP, **1**, **2**, **3**, **4** and **6**) was accomplished according to the method described by Wagner and colleagues starting from the nucleoside monophosphate (77).

Deprotection of the methyl esters was carried out with sodium hydroxide or potassium carbonate in methanol-water solution. Scheme II.1 shows a representative example of such a synthetic scheme using the β-Ala-dAMP analogue as prototype. Deprotection of the benzylated γ-hydroxypropionic acid-dAMP was easily achieved by hydrogenation.

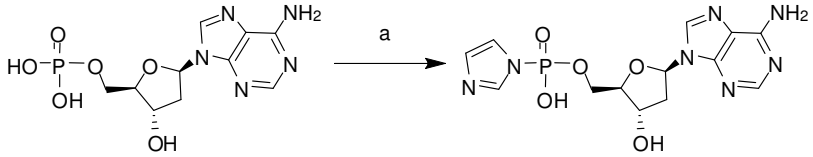



**Scheme II.1**. Reaction scheme for dicyclohexylcarbodiimide assisted coupling and methyl ester deprotection: a) DCC, *tert*butanol, water, reflux, 6-8 hrs; b) R=Me: sodium hydroxide, methanol, water (1M), rt, 1h; R=Bn: $H_2$ Pd-C, methanol; c) TEAB 1M, 4°C.

For the synthesis of glycolic acid- dAMP **5** and hydroxypropionic acid dAMP **6**, we resorted to a method based on divalent cation assisted coupling (represented for the synthesis of **5** in Scheme II.2), as previously published by Sawai (85). In this approach, the carboxyl acid moiety does not need to be protected, as it is proposed to possibly serve as ligand for the divalent metal-ion during the nucleotidyl transfer, thus reducing the number of synthetic steps to one. The starting material is the activated (deoxy)nucleotide dAMP imidazolide (ImdAMP), synthesized by adapting the method of Lohrman and Orgel (29), *via* a Mukayiama oxidation-reduction condensation (Scheme II.3) (31;86).

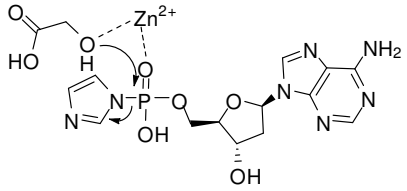



**Scheme II.2:** Reaction scheme for divalent cation mediated coupling of glycolic acid and dAMP: a) zinc chloride, *N*-ethyl morpholine buffer pH 7.5, rt, 2-4 days.



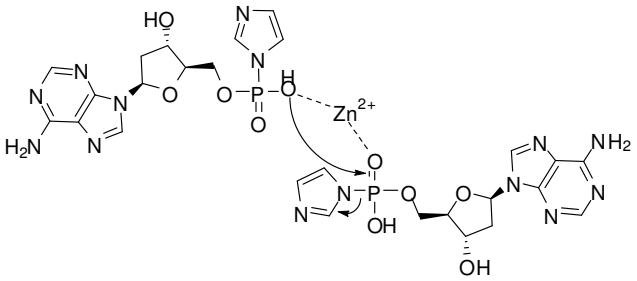

**Scheme II.3:** Synthesis of ImdAMP by oxidation-reduction condensation: a) Triphenylphosphine, imidazole, dipyridyl disulfide, triethylamine, DMSO, rt, 30 min.

In this mechanism the divalent metal ion is thought to be activating the phosphate group by chelation and also chelating the carboxylic acid function, therefore bringing the nucleophilic hydroxyl group in proximity for in-line attack of the phosphorous, as shown in Scheme II.4.



**Scheme II.4:** Mechanism of divalent cation assisted coupling as proposed by Sawai (85).

$^{31}$P-NMR monitoring of two to four day-reaction showed disappearance of the ImdAMP peak (-8 ppm) and formation of the phosphodiester (1.4 ppm), concomitant with the appearance of an additional peak at -11 ppm. This side product of the reaction was purified and characterized as bis(2′-nucleoside) 5′,5′-diphosphate (dAppdA). Indeed, the yield of this coupling reaction is limited by the simultaneous dimerization of ImdAMP to the pyrophosphate dimer dAppdA. It is possible that two imidazolide molecules, bound in a single $Zn^{2+}$ complex, react together to form an inter-nucleotide bond (30), as shown in Scheme II.5. Such dimerization in concentrated solutions of phosphoimidazolide nucleotide was also observed in the presence of magnesium ions.



**Scheme II.5:** Possible mechanism for the formation of dAppdA starting from ImpdA.

The ratio between desired phosphodiester and side product dAppdA depends on the metal species, as shown in Figure II.1. In the presence of zinc, the reaction yields 80-90 % of the desired product. Presence of divalent magnesium ion promotes dAppdA formation and only yields about 40 % of phosphodiester.

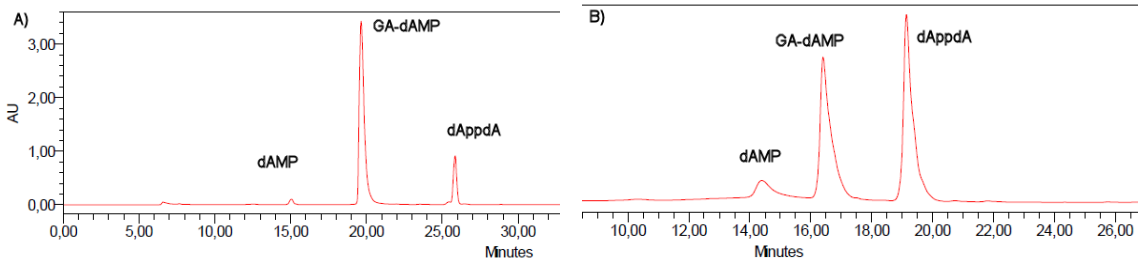



**Figure II.1:** Reverse phase HPLC purification profiles of GA-dAMP synthesis in presence of A) zinc chloride, B) magnesium chloride after 24 h reaction.

## II.2.2 Single incorporation experiments

Previous experiments carried out with the L-aspartic acid phosphoramidate of dAMP demonstrated that this amino acid was an acceptable leaving group for the polymerase in the nucleotidyl insertion process. Incorporation results were lower when using the enantiomeric D-aspartic acid leaving group.

In the present study, we evaluated the capacity to incorporate a deoxyadenosine nucleotide in the primer-template complex P1:T1 (Table II.1) using HIV-1 RT and the above mentioned dATP analogues as substrates. The initial screening was carried out using a template with an overhang of one thymidine nucleotide followed by three deoxycytidine nucleotides. Incorporation efficiency was analyzed by the polyacrylamide gel-based single nucleotide incorporation assay (87;88).

**Table II.1.** Primer-template complex used in the incorporation assay with HIV-1 RT DNA polymerase. Bold letters indicate the template overhang in the hybridized primer-template duplex.

| Oligonucleotide | Sequence |
|---|---|
| P1 | 5'-CAGGAAACAGCTATGAC-3' |
| T1 | 3'-GTCCTTTGTCGATACTG**TCCC**-5' |

The isophtalic acid phosphodiester analogue **1** is recognised by HIV-1 RT and efficiently incorporated into a growing primer strand with a conversion to a (P+1) strand in 90-92 % yield over a period of 2 hours at 1mM concentration (figure II.2). The corresponding aniline analogue **2** is less well recognized as substrate (67% (P+1) product). Finally, little incorporation (13% (P+1) product after 2 hours) was observed using the phtalic acid dAMP derivative **3**.

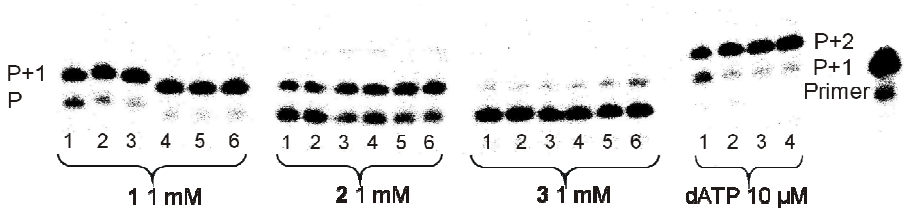



**Figure II.2:** Profiles of incorporation of dAMP into P1:T1 by HIV-1 reverse transcriptase using **1** (5-OH-*i*PA)-, **2** (5-NH-*i*PA)-, and **3** (4-OH-PA)-dAMP as substrates. Aliquots were drawn at 5, 10, 20, 30, 60 and 120 minutes (for dATP: 5, 30, 60, 120 min).

A few interesting observations can be made from this first panel. Despite the geometric constraint brought by the aromatic ring, dicarboxylic acid substituted phenols and aniline can still function as leaving group in the polymerase catalyzed reaction. As expected, a phenolate is a better leaving group than the amide ion since it is a better 'base' although it is not clear if protonation at the nitrogen atom of the aniline group is involved in the catalytic mechanism. It has been shown recently that protonation of the pyrophosphate leaving group is important and occurs thanks to the presence of a neighbouring acid residue of the active site (89).

Among the two phenol moieties, the one carrying both substituents in *meta*-position (**1**) is the most successful PPi mimic. This could indicate that the orientation of both carboxyl functions is important, which might be attributed to steric hindrance in the active site of the polymerase, or to a more stable complex. Compound **1** was further evaluated at different concentrations

(Figure II.3). At 500 µM it displayed a maximum of 75% of (P+1) formation, which represents 88% of the yield obtained with L-Asp-dAMP as a substrate (81).

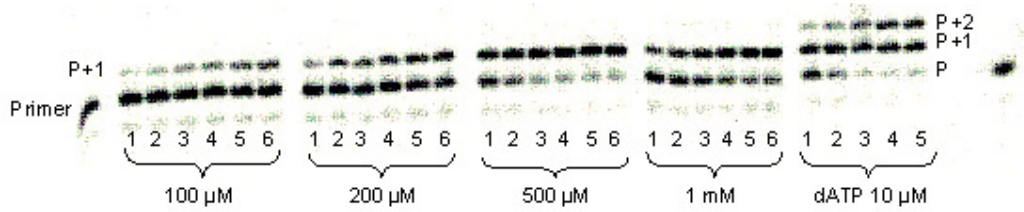



**Figure II.3:** Profiles of incorporation of dAMP into P1:T1 with dAMP by HIV-1 reverse transcriptase using 5-OH-*i*PA-dAMP **1** as substrate. Aliquots were drawn at 5, 10, 20, 30, 60 and 120 minutes (for dATP: 5, 10, 20, 30, 60 min).

These leaving groups could also only serve as activating groups in the reaction medium and be incorporated in a polymerase independent manner, as dAMP imidazolide (Im-dAMP) is incorporated into an appropriate primer:template duplex in a polymerase-independent reaction (90). Therefore, a control was made in the absence of polymerase, and we could observe that no primer elongation takes place in these conditions.

Using the oxygen analogue of glycine, i.e. glycolic acid deoxyadenosine monophosphate phosphodiester **5** (GA-dAMP), the incorporation efficiency is better than with glycine itself. The compound possesses a reduced chelating capacity when compared with L-Asp, which apparently did not prevent incorporation of one nucleotide unit. The incorporation yield is moderate, as (P+1) formation was observed up to 68% at a concentration of 500 µM (Figure II.4). A large difference between the glycine phosphoramidate analogue (13.8 % (P+1) formation) and its phosphodiester counterpart could be observed, showing the better leaving group properties of the phosphodiester candidate.

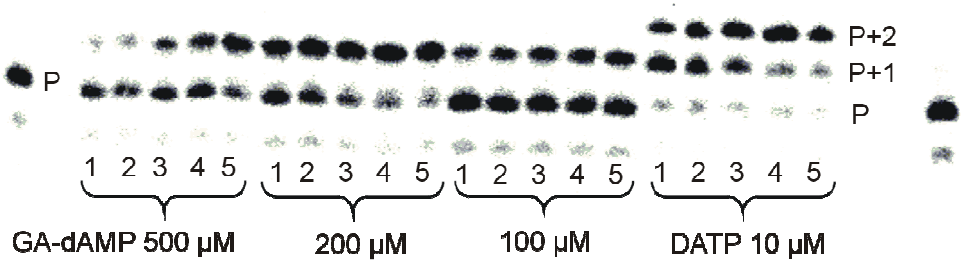



**Figure II.4:** Profile of incorporation of dAMP into P1:T1 with dAMP by HIV-1 reverse transcriptase using GA-dAMP **5** into P1:T1 as substrate. Aliquots were drawn at 5, 10, 20, 30 and 60 minutes.

Interestingly, dAppdA (side product from the synthesis of GA-dAMP) also serves a substrate for HIV-1 RT. This bulky substrate is recognized by HIV-1 RT and dAMP is incorporated into dsDNA, resulting in up to 95% of extended primer (Figure II.5). Possible contamination of the GA-dAMP solution by the dimer was therefore carefully monitored by available spectroscopic methods.

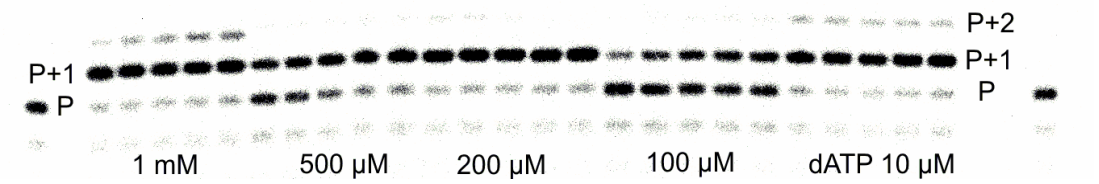



**Figure II.5:** Profile of incorporation of dAMP into P1:T1 with dAMP by HIV-1 reverse transcriptase using dAppdA as substrate. Aliquots were drawn at 5, 10, 20, 30 and 60 minutes.

The phosphoramidate analogue aminopropionic acid dAMP **4,** whose alkyl chain contains one additional methylene group, is not a substrate for HIV-1 RT, as no (P+1) product formation was observed at four concentrations ranging from 1 mM to 100 µM. On the contrary, the phosphodiester counterpart hydroxypropionic acid dAMP **6** serves as substrate, albeit poorly; (P+1) product formation was observed up to 11% at 500 µM concentration. Table II.2 gives an overview of the yield of dAMP incorporation (%) into DNA (at 60 minutes incubation time) using substrates **1**, **2**, **4-6**.

**Table II.2.** Percentage of incorporation of dAMP in function of the leaving group moiety and concentration (60 min incubation time, P1:T1, HIV-1 RT).

| | *i*PA | | -$CH_2$-COOH | | -$CH_2$-$CH_2$-COOH | |
|---|---|---|---|---|---|---|
| | O-P (1) | N-P (2) | O-P (5) | N-P | O-P (6) | N-P (4) |
| 100 µM | 24 | nd | 41 | nd | 0 | 0 |
| 200 µM | 47 | nd | 78 | nd | 0 | 0 |
| 500 µM | 75 | nd | 67 | 20 | 5 | 0 |
| 1 mM | 87 | 13 | nd | nd | 11 | 0 |

[a] % of the total amount of radioemitting oligonucleotides in the mixture

From Table II.2 we can conclude that a phosphodiester (P-O) bond is cleaved easier than a phosphoramidate (P-N) bond. Surprisingly, in the case of a phosphodiester bond between the

nucleoside monophosphate and the leaving group, one carboxylic acid function on the leaving group is sufficient to enable the nucleotidyl transfer reaction. This is not the case when considering the phosphoramidate linkage. Incorporation studies with aminopropionic acid dAMP and Gly-dAMP demonstrate that phosphoramidate bond (of an aliphatic leaving group) can only be cleaved when two carboxylate groups are present. In addition, the difference in incorporation yield between **5** and **6** suggest that the closer the carboxylic acid is, the better the group mimics pyrophosphate.

## *II.3 Discussion*

The propagation of new information systems *in vivo* for synthetic biology purposes requires that the system is orthogonal to the existing natural informational systems (DNA and RNA). Only in this case, it can be avoided that the natural system will become infiltrated by information from outside. This also applies to the precursors for the enzymatic synthesis of artificial nucleic acids, which means that modified building blocks used for the *in vivo* synthesis of artificial nucleic acids should not enter the cellular metabolic pathways, since it may lead to toxicity. One way to realize this is to develop an independent metabolic route for chemically modified polymerase substrates. This might be based on the selection of new leaving groups for polymerase-catalyzed nucleic acid synthesis. Preferentially, this leaving group is metabolically available and can be recycled after the reaction.

Deoxyadenosine triphosphate analogues where the pyrophosphate moiety is replaced by L-aspartic acid had previously shown to act as substrates for DNA polymerization catalyzed by HIV-1 RT. In the presence of only one carboxylic acid in the leaving group, incorporation is observed when a phosphodiester linkage is cleaved during the reaction and not when a phosphoramidate bond has to be cleaved. In this case the better leaving group capacity of an oxygen linked moiety could compensate for the loss in electrostatic effects. This observation is confirmed when using leaving groups with an aromatic character. The difference between phtalic acid and isophtalic acid is interesting, and is most probably related to a different way that these molecules can be accommodated in the active site of the polymerase. Although the results do not surpass those found for L-Asp-dAMP, they contribute to the study of the structure-substrate relationship of HIV-1 RT and demonstrate that a variety of leaving groups can be used to synthesize DNA in an enzymatic way in place of pyrophosphate.

## *II.4 Experimental section*

### II.4.1. General methods

NMR spectral analyses were carried out on Brucker Avance™ II 300 MHz or 500 MHz with PAXTI probe and recorded at 25°C, unless otherwise specified. The Bruker TopspinTM 2.1 software was used to process spectra. Chemical shifts are expressed in parts per million (ppm) by frequency. $^{1}H$- and $^{13}C$- NMR chemical shifts are referenced to an internal TMS signal (δ = 0.00 ppm), $^{31}P$ NMR chemical shifts are referenced to an external 85 % $H_3PO_4$ standard (δ = 0.00 ppm). Standard mass spectra were measured with a Finnigan LCQ DuO (Thermo Fischer Scientific) using the electrospray ionization technique (ESI); data were acquired with the LAC/E32 system (Waters). Exact mass spectra were obtained with a Q-Tof 2TM (Micromass Ltd.) coupled to a CapLC™ system (Waters). Chemicals of analytical or synthetic grade were obtained from commercial sources and were used as received (deoxyadenosine monophosphate: Sigma Aldrich; palladium on carbon: Aldrich; DCC, glycolic acid and zinc chloride: Fluka; *N*-ethylmorpholine, phtalic acids, *tert*-butanol, thionyl chloride and triethylamine: Acros). Dimethyl or benzyl ester of dicarboxylic acids and were either obtained form commercial sources or prepared by methylation after thionyl chloride activation. Technical solvents were obtained from Brenntag (Deerlijk, Belgium). Acetonitrile HPLC Grade was purchased from Fisher Scientific. Flash silica chromatography was performed on Davisil® silica gel 60, 0.040–0.063 mm (Grace Davison). Thin Layer Chromatography was performed on Alugram® silica gel UV254 mesh 60, 0.20 mm (Macherey-Nagel). Celite® 545 was purchased at VWR international.

### II.4.2 General procedure for the synthesis of the ester intermediates

***Synthesis of compounds 1 - 4***

The example described is for 2'-deoxyadenosine-5'-(5-amino-isophtalic acid dimethyl ester) phosphoramidate (intermediate for 2). In a two-neck flask, 2'-deoxyadenosine-5'-O-monophosphoric acid hydrate (50 mg, 0.14 mmoles), 5-amino-isophtalic acid dimethyl ester hydrochloride (245 mg, 1 mmoles, 7 equiv.) and dicyclohexylcarbodiimide (DCC) (147 mg, 1 mmol, 7 equiv.) were dissolved under argon in a mixture of *tert*-butanol (3mL) and $H_2O$ (1mL). A few drops of triethylamine ($Et_3N$) were added to the solution as hydrochloric acid scavenger. The reaction mixture was refluxed for 6 hours while stirring under argon. The progress of the reaction was monitored by TLC (*i*PrOH:$H_2O$:$NH_3$ 7:2:1). The reaction mixture

was cooled and the solvent was removed by rotary evaporation (37°C). The product was isolated by silica column chromatography eluting with $CHCl_3$:MeOH:$H_2O$ gradient (5:1:0, 5:2:0.25, 5:3:0.5) affording the product as a white solid (35 mg, 47% yield). $^1$H NMR (300 MHz, $D_2O$): δ 7.97 (s, 1H, $H_8$), 7.94 (s, 1H, $H_{iPA}$), 7.61 (s, 1H, $H_2$), 7.44 (s, 2H, $H_{iPA}$), 6.25 (apparent t, $^3J_{H1'-H2'}$ = 6.6 Hz, 1H, $H_{1'}$), 4.67 (m, 1H, $H_{3'}$), 4.32 (m, 1H, $H_{4'}$), 4.08 (m, 2H, $H_{5'}$), 3.9 (s, 6H, $CH_3$), 2.85 (m, 1H, $H_{2'}$), 2.53 (m, 1H, $H_{2''}$) ppm. $^{13}$C NMR (75 MHz, $D_2O$): δ 167.8, 154.6, 151 .8, 147.9, 142.4, 139.7, 129.6, 121.7, 121.6, 121.1, 86.1, 83.9, 71.1, 65.2, 58.8, 52.7, 46.6, 38.2 ppm. $^{31}$P NMR (121 MHz, $D_2O$): δ -3.50 ppm. MS (ESI, negative ion mode): m/z calculated for $C_{20}H_{23}N_6O_9P$: 522.4, found 521.8.

***Deprotection of the ester intermediate. The example described is for the synthesis of 2'-deoxyadenosine-5'-(5-aminoisophtalic acid) phosphoramidate (2).***

A solution (1 mL) of 1.3 M potassium carbonate ($K_2CO_3$) in MeOH:$H_2O$ 2:1 was added to 2'-deoxyadenine-5'-(5-aminoisophtalic acid dimethyl ester) phosphoramidate (35 mg, 67 μmoles) and the reaction was carried out at room temperature while stirring under argon for 4 hours. The course of the reaction was monitored by TLC *i*PrOH:$H_2O$:$NH_3$ 7:2:1). Once the starting material has disappeared, the reaction mixture was neutralized by addition of 2M triethylammonium bicarbonate. The solvent was removed under reduced pressure and the resulting residue was dried by lyophilisation. The desired product was purified by silica column chromatography eluting with (*i*PrOH:$H_2O$:$NH_3$) gradient. The product was isolated and concentrated by lyophilisation to provide a white solid (20 mg, 60%). $^1$H NMR (300 MHz, $D_2O$): $\delta$ 8.08 (s, 1H, $H_8$), 8.00 (s, 1H, $H_2$), 7.59 (s, 1H, $H_{iPA}$), 7.31 (s, 2H, $H_{iPA}$), 6.29 (apparent t, $^3J_{H1'-H2'}$ = 6.6 Hz, 1H, $H_{1'}$), 4.55 (m, 1H, $H_{3'}$), 4.22 (m, 1H, $H_{4'}$), 4.00 (m, 2H, $H_{5'}$), 2.80 (m, 1H, $H_{2'}$), 2.45 (m, 1H, $H_{2''}$) ppm. $^{13}$C NMR (75 MHz, $D_2O$): $\delta$ 172.7, 153.7, 150.4, 148.0, 141.4, 140.0, 135.1, 121.0, 120.0, 119.9, 85.9, 84.0, 71.0, 64.3, 38.4 ppm. $^{31}$P NMR (121 MHz, $D_2O$): $\delta$ - 0.83 ppm. High res. MS (ESI, negative ion mode): m/z calculated for $C_{18}H_{19}N_6O_9P$: 494.09511, found: 493.08740. Stability in $D_2O$ at r.t.: superior as 24h.

***2'-deoxyadenosine-5'-(5-hydroxy-isophtalic acid dimethyl ester) phosphodiester (intermediate for 1)***

Yield: 62 mg (79%). $^1$H NMR (300 MHz, $D_2O$): $\delta$ 8.23 (s, 1H, $H_8$), 8.06 (s, 1H, $H_2$), 8.00 (s, 1H, $H_{iPA}$) 7.68 (s, 2H, $H_{iPA}$), 6.40 (apparent t, $^3J_{H1'-H2'}$ = 6.7 Hz, 1H, $H_{1'}$), 4.4 (m, 1H, $H_{3'}$), 4.27 (m, 1H, $H_{4'}$), 4.00 (m, 2H, $H_{5'}$), 3.9 (s, 3H, $CH_3$), 3.8 (s, 3H, $CH_3$), 2.90 (m, 1H, $H_{2'}$),

2.56 (m, 1H, $H_{2'}$) ppm. $^{13}C$ NMR (75 MHz, $D_2O$): $\delta$ 166.9, 151.8, 151.6, 151.6, 139.9, 139.4, 130.6, 125.4, 125.3, 118.8, 85.9, 83.9, 71.2, 65.9, 52.9, 38.2 ppm. $^{31}P$ NMR (121 MHz, $D_2O$): $\delta$ - 5.05 ppm. MS (ESI negative ion mode): m/z calculated for $C_{20}H_{22}N_5O_{10}P$: 523.11043, found: 522.1035.

***2'-deoxyadenosine-5'-(5-hydroxy-isophtalic acid) phosphodiester (1)***

Yield: 25 mg (66 %). $^1H$ NMR (300 MHz, $D_2O$): $\delta$ 8.23 (s, 1H, $H_8$), 8.1 (s, 1H, $H_2$), 7.7 (s, 1H, $H_{iPA}$), 7.6 (s, 2H, $H_{iPA}$), 6.46 (apparent t, $^3J_{H1'-H2'}$ = 6.7 Hz, 1H, $H_{1'}$), 4.65 (m, 1H, $H_{3'}$), 4.32 (m, 1H, $H_{4'}$), 4.19 (t, 2H, $H_{5'}$), 2.84 (m, 1H, $H_{2'}$), 2.56 (m, 1H, $_{H2'}$) ppm. $^{13}C$ NMR (75 MHz, $D_2O$): $\delta$ 173.7, 155.0, 152.0, 151.8, 149.0, 140.6, 137.5, 125.3, 123.4, 119.0, 96.5, 84.7, 71.8, 66.1, 38.8 ppm. $^{31}P$ NMR (121 MHz, $D_2O$): $\delta$ - 4.4 ppm. High res. MS (ESI, negative ion mode): m/z calculated for $C_{18}H_{19}N_5O_{10}P$: 495.07913 found: 494.0716. Stability in $D_2O$ at r.t.: superior as 24h.

***2'-deoxyadenosine-5'-(4-hydroxy-phtalic acid dimethyl ester) phosphodiester (intermediate for 3)***

Yield: 70 mg (93%). $^1H$ NMR (300 MHz, $D_2O$): $\delta$ 8.12 (s, 1H, $H_8$), 8.09, (s, 1H, $H_2$), 7.48 (d, 1H, $H_{PA}$), 7.16 (d, 1H, $H_{PA}$), 7.13 (s, 1H, $H_{PA}$), 6.4 (apparent t, $^3J_{H1'-H2'}$ = 6.8 Hz, 1H, $H_{1'}$), 4.70 (m, 1H, $H_{3'}$), 4.31 (m, 1H, $H_{4'}$), 4.21 (m, 2H, $H_{5'}$), 3.9 (s, 3H, $CH_3$), 3.8 (s, 3H, $CH_3$), 2.81 (m, 1H, $H_{2'}$), 2.54 (m, 1H, $H_{2'}$) ppm. $^{13}C$ NMR (75 MHz, $D_2O$): $\delta$ 169.4, 168.5, 154.3, 154.2, 150.4, 148.2, 140.2, 133.1, 130.8, 124.7, 122.6, 119.9, 118.4, 85.9, 84.0, 71.0, 65.8, 53.3, 53.1, 38.5 ppm. $^{31}P$ NMR (121 MHz, $D_2O$): $\delta$ -5.42 ppm. MS (ESI negative ion mode): m/z calculated for $C_{20}H_{22}N_5O_{10}P$: 523.4, found 522.3.

***2'-deoxyadenosine-5'-(4-hydroxy-phtalic acid) phosphodiester (3)***

Yield: 25 mg (66%). $^1H$ NMR (300 MHz, $D_2O$): $\delta$ 8.32 (s, 1H, $H_8$), 8.21 (s, 1H, $H_2$), 7.33 (d, 1H, HPA), 7.12 (s, 1H, HPA), 6.92 (d, 1H, HPA), 6.49 (apparent t, $^3J_{H1'-H2'}$ = 6.7 Hz, 1H, $H_{1'}$), 4.68 (m, 1H, $H_{3'}$), 4.30 (m, 1H, $H_{4'}$), 4.17 (m, 2H, $H_{5'}$), 2.88 (m, 1H, $H_{2'}$), 2.61 (m, 1H, $H_{2'}$) ppm. $^{13}C$ NMR (125 MHz, $D_2O$): $\delta$ 181.3, 178.3, 175.8, 175.2, 155.0, 152.1, 151.8, 148.4, 139.5, 131.5, 129.2, 119.2, 118.3, 85.6, 83.9, 71.2, 65.7, 38.6 ppm. $^{31}P$ NMR (121 MHz, $D_2O$): $\delta$ -4.6 ppm. High res. MS (ESI positive ion mode): m/z calculated for $C_{18}H_{19}N_6O_9P$: 496.0864, found: 496.8296. Stability in $D_2O$ at r.t.: superior as 24h.

***2'-deoxyadenosine-5'- β-alanine phosphoramidate methyl ester (intermediate for 4)***

Yield: 172 mg (81%). $^{1}$H NMR (500 MHz, $CD_3OD$): $\delta$ 8.5 (s, 1H, $H_8$); 8.2 (s, 1H, $H_2$); 6.5 (apparent t, $^{3}J_{H1'-H2'}$ = 7.0 Hz, 1H, $H_{1'}$); 4.7 (m, 1H, $H_{3'}$); 4.2 (m, 1H, $H_{4'}$); 4.08-3.98 (m, 2H, $H_{5'}$); 3.64 (s, 3H, -O-$CH_3$); 3.14-3.10 (m, 2H, $CH_{2\beta}$); 2.84-2.80 (m, 1H, $H_{2'}$); 2.47-2.45 (m, 3H, H2′, $CH_{2\alpha}$) ppm. $^{13}$C NMR (125 MHz, $CD_3OD$): $\delta$ 173.3, 155.9, 152.4, 149.0, 139.7, 118.7, 86.6, 83.9, 71.7, 64.1, 50.6, 39.9, 37.2, 35.9 ppm. $^{31}$P NMR (202 MHz, $CD_3OD$): $\delta$ 7.53 ppm. High res. MS (ESI negative ion mode): m/z calculated for $C_{14}H_{21}N_6O_7P$: 416.1209, found: 415.1127.

***2'-deoxyadenosine-5'- β-alanine phosphoramidate (4)***

Yield: 15 mg (87%). $^{1}$H NMR (500 MHz, $D_2O$): $\delta$ 8.2 (s, 1H, $H_8$), 7.8 (s, 1H, $H_2$), 6.2 (apparent t, $^{3}J_{H1'-H2'}$ = 6.7 Hz, 1H, $H_{1'}$), 4.6 (m, 1H, $H_{3'}$), 4.1 (m, 1H, $H_{4'}$), 3.8 (m, 2H, $H_{5'}$), 2.8 (m, 2H, $CH_{2\beta}$), 2.7 (m, 1H, $H_{2'}$), 2.5 (m, 1H, $H_{2'}$), 2.1 (t, $J_{H\alpha-H\beta}$ = 6.0 Hz, 2H, $CH_{2\alpha}$) ppm. $^{13}$C NMR (125 MHz, $D_2O$): $\delta$ 178.1, 154.3, 151.0, 148.3, 140.0, 118.7, 85.7, 83.5, 70.9, 63.6, 38.5 (2C), 37.1 ppm. $^{31}$P NMR (202 MHz, $D_2O$): $\delta$ 8.75 ppm. High res. MS (ESI negative ion mode): m/z calculated for $C_{13}H_{19}N_6O_7P$: 402.10528, found: 401.0977. Stability in $D_2O$ at r.t.: superior as 24h.

***2'-deoxyadenosine-5'-O-monophosphate imidazolate***

A mixture of dAMP (100 mg, 0.3 mmoles), dithiopyridine (211 mg, 0.96 mmoles 3.2 eq.), triphenylphosphine (327 mg, 0.96 mmoles, 3.2 eq.) and imidazole (327 mg, 4.8 mmoles, 16 eq.) was dried under high vacuum for 30 minutes. Subsequent dissolution in anhydrous DMSO (3.5 mL) under argon atmosphere afforded a clear yellow solution. Triethylamine (95 µL, 0.7 mmoles, 2.3 eq.) was added *via* a syringe and the resulting solution stirred at room temperature for 4 hours. It was then dropped in a -20°C 0.1M sodium iodide solution in dry acetone and allowed to precipitate for 15 minutes. Filtration of the resulting suspension and repeated washing thereof with cold dry acetone afforded a white solid. Further drying on $POCl_3$ under HV afforded quantitative yield of dAMP-imidazolide (110 mg, 96%).

$^{1}$H NMR (300 MHz, $D_2O$, 5°C): $\delta$ 8.21 (s, 1H, $H_8$), 8.20 (s, 1H, $H_2$), 7.7 (s, 1H, $H_{Im}$), 7.0 (s, 1H, $H_{Im}$), 6.8 (s, 1H, $H_{Im}$), 6.4 (apparent t, $^{3}J_{H1'-H2'}$ = 6.41 Hz, 1H, $H_{1'}$), 4.6 (m, 1H, $H_{3'}$), 4.2 (m, 1H, $H_{4'}$), 4.0 (m, 2H, $H_{5'}$), 2.9 (m, 1H, $H_{2'}$), 2.6 (m, 1H, $H_{2'}$) ppm. $^{31}$P NMR (121 MHz,

$D_2O$): $\delta$ -8.01 ppm. High res. MS (ESI negative ion mode): calculated for $C_{13}H_{16}N_7O_5P$: 381.0951, found: 380.0869. Stability in $D_2O$ at r.t.: 10% hydrolysis after 12h.

***2'-deoxyadenosine-5'-(glycolic acid) phosphodiester (5)***

A solution of 2'-deoxyadenosine-5'-O-monophosphate imidazolate (20 mg, 50 µmoles), hydroxyacetic acid (7.6 mg, 100 µmole, 2 equiv.), zinc chloride (7 mg, 50 µmoles, 1 equiv.) in *N*-ethylmorpholine aqueous buffer 0.2 M (2 mL, pH 7.5) was stirred under argon at room temperature for 1 day. The reaction was monitored by TLC (*i*PrOH:$H_2O$:$NH_3$ 7:2:1), quenched with an EDTA (ethylenediaminetetraacetic acid) 0.25 M solution in order to break-down the nucleotide-metal complex, and finally lyophilized affording a white solid (42%). The product was isolated after HPLC purification on a PLRP-S column (100 Å 8 µm, 300x7.5 mm, Polymer Laboratories) running a gradient of acetonitrile in 50 mM triethylammonium acetate buffer. Yield 41% (from HPLC purification profile). $^1$H NMR (300 MHz, $D_2O$, 5°C): $\delta$ 8.4 (s, 1H, $H_8$), 8.1 (s, 1H, $H_2$), 6.4 (apparent t, $^3J_{H1'-H2'}$ = 6.8 Hz, 1H, $H_{1'}$), 4.5 (m, 1H, $H_{3'}$), 4.2 (m, 1H, $H_{4'}$), 4.1 (s, 2H, $CH_2$), 3.9 (m, 2H, $H_{5'}$), 2.8 (m, 2H, $CH_{2\alpha}$), 2.7 (m, 1H, $H_{2'}$), 2.5 (m, 1H, $H_{2'}$) ppm. $^{13}$C NMR (125 MHz, $D_2O$): $\delta$ 180.3, 155.8, 152.8, 148.93, 140.22, 118.8, 86.5, 84.0, 71.8, 64.6, 54.1, 39.5 ppm. $^{31}$P NMR (121 MHz, $D_2O$): $\delta$ 1.44 ppm. High res. MS (ESI negative ion mode): m/z calculated for $C_{12}H_{16}N_5O_8P$: 389.07365, found: 388.0652. Stability in $D_2O$ at r.t.: 12h. Unstable in RTRB.

***5',5'-Diphosphate-2',2'-dideoxyadenosine (dAppdA)***

Purified on C18 HPLC (Alltima® C18 column (5µm, 250x10mm, GRACE) eluted at 25% TEAA 50mm. $^1$H NMR (300 MHz, $D_2O$, 25°C): $\delta$ 8.07 (s, 1H, $H_8$), 7.92 (s, 1H, $H_2$), 6.20 (apparent t, $^3J_{H1'-H2'}$ = 6.21 Hz, 1H, $H_{1'}$), 4.55 (m, 1H, $H_{3'}$), 4.13 (m, 3H, $H_{4'}$&$H_{5'}$), 2.50 (m, 1H, $H_{2'}$), 2.43 (m, 1H, $H_{2'}$) ppm. $^{13}$C NMR (125 MHz, $D_2O$): $\delta$ 155.06, 152.64, 148.36, 140.06, 118.40, 86.52, 84.35, 71.24, 66.00, 40.04 ppm. $^{31}$P NMR (121 MHz, $D_2O$): $\delta$ -11.23 ppm. HRMS (ESI, negative ion mode): m/z calculated for $C_{20}H_{25}N_{10}O_{11}P_2^-$: 643.1185, found: 643.1183. Stable in $D_2O$ at r.t.

***2'-deoxyadenosine-5'-(γ-hydroxypropionic acid benzyl ester) phosphodiester (intermediate for 6)***

Yield: 94 mg (68%). $^1$H NMR (300 MHz, $D_2O$): δ 8.21 (s, 1H, $H_8$), 8.02 (s, 1H, $H_2$), 7.12 (m, 5H, Ph), 6.29 (m, 1H, $H_{1'}$), 4.91 (s, 2H, $PhCH_2$), 4.55 (m, 1H, $H_{3'}$), 4.12 (m, 1H, $H_{4'}$), 4.95 (m,

2H, $H_{5'}$), 3.87 (m, 2H, γ-$CH_2$), 2.55 (m, 3H, $H_{2'a}$ and β-$CH_2$), 2.44 (m, 1H, $H_{2'b}$). $^{13}$C NMR (75 MHz, $D_2O$): $\delta$ 173.22, 155.32, 152.48, 148.41, 139.48, 135.13, 128.40 (2C), 128.03, 127.27 (2C), 118.52, 85.68, 83.69, 71.05, 66.69, 65.05, 61.43, 39.11, 35.42 ppm. $^{31}$P NMR (121 MHz, $D_2O$): $\delta$ 0.029 ppm. High res. MS (ESI, negative ion mode): m/z calculated for $C_{20}H_{24}N_5O_8P$: 493.1363, found: 492.1281.

***2′-deoxyadenosine-5′-(γ-hydroxypropionic acid) phosphodiester (6)***

A mixture of 2′-deoxyadenosine-5’-(3-hydroxypropionic acid benzyl ester) phosphodiester (40 mg, 0.0811 mmoles) and 10% palladium on activated carbon (35 mg, 0.0331 mmoles) in methanol (10 mL) was stirred under hydrogen atmosphere at room temperature for 1.5h. The progress of the reaction was monitored by TLC (*i*PrOH:$H_2O$:$NH_3$ 7:1:2). The catalyst was filtered off on Celite® and the filtrate was concentrated to dryness in vacuum. The residue was redissolved in $H_2O$, and lyophilized affording the product as a white solid (29 mg, 89% yield). $^1$H NMR (300 MHz, $D_2O$): $\delta$ 8.34 (s, 1H), 8.15 (s, 1H), 6.42 (apparent t, 1H, $^3J_{H1'-H2'}$ = 6.78Hz, $H_{1'}$), 4.64 (m, 1H, $H_{3'}$), 4.19 (m, 1H, $H_{4'}$), 3.94 (m, 2H, $H_{5'}$), 3.86 (dd, 2H, $^3J_{\gamma-\beta}$ = 6.4Hz, $^2J_{\gamma1-\gamma2}$ = 12.78Hz, γ-$CH_2O$), 2.75 (m, 1H, $H_{2'a}$), 2.53 (m, 1H, $H_{2'b}$), 2.32 (apparent t, 2H, $^3J_{\gamma-\beta}$ = 6.4Hz, β-$CH_2$) ppm. $^{13}$C NMR (75 MHz, $D_2O$): $\delta$ 179.29, 155.22, 152.34, 148.40, 139.39, 118.27, 85.43, 83.31, 70.81, 64.52, 62.80, 38.56, 38.16 ppm. $^{31}$P NMR (121 MHz, $D_2O$): $\delta$ 0.271 ppm. High res. MS (ESI, negative ion mode): m/z calculated for $C_{13}H_{18}N_5O_8P$: 403.0893 Found: 402.0817. Stability in $D_2O$ at r.t.: superior as 24h. Unstable in RTRB.

## II.4.3 Enzymatic reactions

**DNA duplexes** Oligodeoxyribonucleotides P1 and T1 were purchased from Sigma Genosys. The concentrations were determined with a Varian Cary-300-Bio UV Spectrophotometer. The lyophilized oligonucleotides were dissolved in diethylpyrocarbonate (DEPC)-treated water and stored at -20°C. The primer oligonucleotides were 5'-$^{33}$P-labeled with 5’-[$\gamma^{33}$P]-ATP (Perkin Elmer) using T4 polynucleotide kinase (New England Biolabs) according to the procedure provided by the supplier. The labeled oligonucleotide was further purified using IllustraTM MicrospinTM G-25 columns (GE Healthcare).

**DNA polymerase reactions:** All incorporation and elongation reactions reported in this manuscript were duplicated, unless otherwise specified; tripliquets were realized in case of conflicting results and in order to assess the reproducibility of the results. For the

incorporation of 1, 2, 3, 4, 5 and 6, a series of batch reactions (20 µL) was performed with the enzyme HIV-1 RT (Ambion, 10 U/µL stock solution, specific activity 8.095 U/mg, concentration 1.2 mg/mL). The final mixture contained 125 nM primer template complex, RT buffer (250 mM Tris.HCl, 250 mM KCl, 50 mM $MgCl_2$, 2.5 mM spermidine, 50 mM dithiothreitol; pH 8.3), 0.025 U/µL HIV-1 RT , and different concentrations of phosphoramidate or phosphodiester building blocks (1 mM, 500 µM, 200 µM and 100 µM). In the case of the aromatic analogues 1 and 3, 1 mM was the only concentration tested. In the control reaction with the natural nucleotide, a 10 µM dATP concentration was used. The mixture was incubated at 37°C and aliquots (2.5 µL) were quenched after 5, 10, 20, 30 and 60 min.

**Steady-state kinetics of single nucleotide incorporation:** The steady-state kinetics of single nucleotide incorporation of the iminodiacetate phosphoramidate 1 (IDA-dAMP) and of a natural nucleoside triphosphate (dATP) was determined by gel-based polymerase assay. All kinetic reactions reported in this manuscript were carried out at least five times in order to obtain satisfying reproducibility (for discussion of the reproducibility of the results, see chapter VI). In all the experiments, the template T1 and the primer P1 were used. The primer and template in 1:2 molar ratio were hybridised in a buffer containing 20 mM Tris.HCl, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.3 and used in an amount to provide 125 nM concentration of the primer in each reaction (20 µL). The range of concentrations for iminodiacetate dAMP was optimized according to a $K_M$ value for the incorporation of an individual nucleotide. In the case of HIV-1 RT (Ambion, 10 U/µL), reaction mixtures containing the enzyme in 0.025 U/µL concentration and appropriate substrate concentration to attain 5-25 % conversion (nucleoside phosphoramidate or natural dATP) were incubated at 37°C and run for 8 different time intervals. The reaction aliquots were quenched as described in chapter II.

### II.4.4 Polyacrylamide gel electrophoresis

All polymerase reaction aliquots (2.5 µL) were quenched by the addition of loading buffer (10 µL, 90% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol and 50 mM EDTA). Samples were heated at 85°C for 3-4 minutes prior to analysis by electrophoresis for 2.5 h at 2000 V on a 30 cm x 40 cm x 0.4 mm 20% (19:1 mono:bis) denaturing gel in the presence of a 100 mM Tris-borate, 2.5 mM EDTA buffer (pH 8.3). Products were visualized by

phosphorimaging. The amount of radioactivity in the bands corresponding to the products of enzymatic reactions was determined by using the imaging device Cyclone® and the associated Optiquant® image analysis software (Perkin Elmer).

# III. Chapter three: Influence of the linkage between leaving group and nucleoside on incorporation efficiency

Part of the results reported in this chapter was published in *ChemBioChem* in 2010 under the title 'Influence of the linkage between leaving group and nucleoside on incorporation efficiency in DNA catalyzed by reverse transcriptase': Giraut A., Herdewijn P. (2010), *Chembiochem* **11**, 1399-1403.

## *III.1. Introduction*

Deoxynucleotide incorporation catalyzed by polymerases is a complex event where chemical reactions rates (91), protein conformation dynamics (92), and stability of successive complexes between dsDNA, enzyme and substrate are some of the most important features (17;93). In order to study the mechanism of polymerization, several approaches are possible (3;12). One of them consists in introducing modifications in the natural substrates of polymerases (deox)nucleoside triphosphates ((d)NTP) and analysing their effect on the deoxynucleotide incorporation efficiency (17;84;94). Sucato *et al.* have investigated the influence of the bridging oxygen of the leaving group on nucleotide incorporation by DNA polymerase β (70;71). The bridging oxygen of pyrophosphate (PPi) was replaced by a halogen substituted methylene. The replacement of pyrophosphate by such biphosphonic acids affected the catalytic efficiency and fidelity of nucleotide incorporation by this polymerase. We propose to study the effect of replacing pyrophosphate by structurally different mimics.

In the previous chapter, the qualitative observation was made that the outcome of single nucleotide incorporation reaction by HIV-1 RT is improved when a phosphoramidate linkage is replaced by a phosphodiester linkage. In view of a synthetic biology project to develop orthogonal chemistry for the synthesis of artificial nucleic acids, the pyrophosphate leaving group has been replaced by a completely different molecule. For example, as previously observed, L-aspartic acid-dAMP (**1**) and L-histidine-dAMP (**2**) (Figure III.1) are recognized by HIV-1 RT for nucleotide incorporation: L-Asp-dAMP and L-His-dAMP incorporation results in the formation of 80-90% and 50-60%, respectively, of primer+1 (P+1) product at 0.5 mM substrate concentration and 120 min reaction time. With both substrates however, templated strand elongation of up to 6 nucleotides was only moderately successful (82).

Pausing and 90% of termination of DNA synthesis were observed after incorporation of two or three nucleotides for both L-His- and L-Asp-dAMP. The study presented in chapter two performed with non-amino-acid-based analogues showed that a phosphodiester linkage was easier to cleave by the enzyme than its phosphoramidate counterpart. Therefore we proposed to synthesize the corresponding phosphodiesters, i.e. L-malic acid- and L-β-imidazole lactic acid- deoxyadenosine monophosphates (L-MA-dAMP **3** and L-ILA-dAMP **4**, resp., Figure III.1).

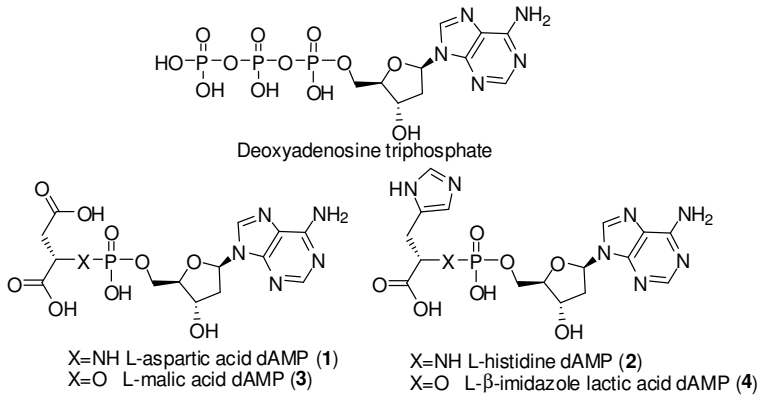



**Figure III.1:** Chemical structures of dATP, L-Asp-dAMP **1**, L-malic acid dAMP **3**, L-histidine-dAMP **2** and L-β-imidazole-lactic acid dAMP **4**.

## *III.2. Results*

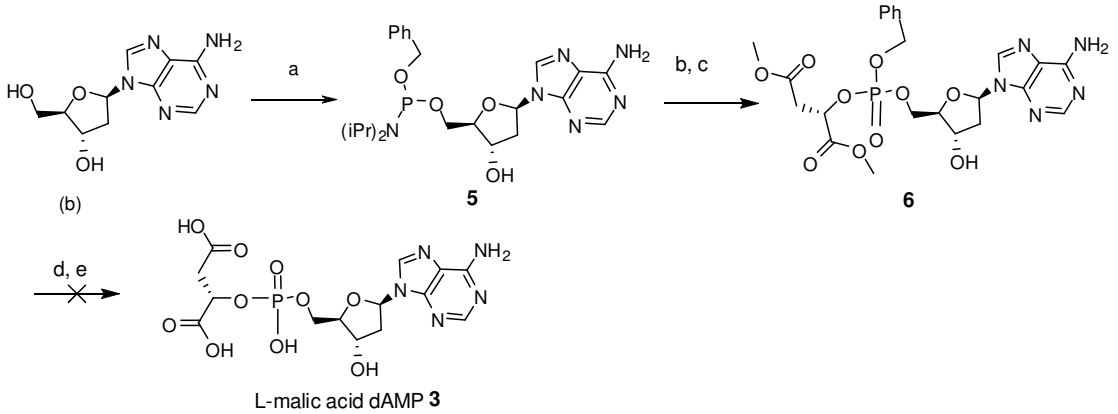
### III.2.1 Synthesis of the phosphoramidate analogue

The synthesis of the phosphoramidate L-His-dAMP was performed as described by Adelfinskaya *et al.* following a dicyclohexylcarbodiimide coupling procedure and subsequent deprotection of the carboxylic acid function in basic conditions (81).

### III.2.2 Synthesis of the phosphodiester analogues

Diverse phosphate coupling methods were investigated for this purpose, starting with the straightforward and good yielding phosphite triester or phosphoramidite method (95). This methodology mostly makes use of cyanoethyl as protecting group for the phosphate (96;97), final phosphate deprotection being carried out in concentrated ammonia. In order to be able to carry out the final deprotection in mild conditions we opted for benzyl as phosphate protecting group. In a preliminary approach carried out by a former member of the laboratory

(shown in scheme III.1), deoxyadenosine was phosphorylated with a diisopropylamino substituted, benzyl protected phosphorous chloride in the presence of DIPEA(98), resulting in the formation of a phosphoramidite (intermediate **5**). Activation of the phosphate by tetrazole allowed nucleophilic attack by the hydroxyl group of malic acid, oxidation with iodine in the presence of water and pyridine yielded the desired phosphotriester **6**.

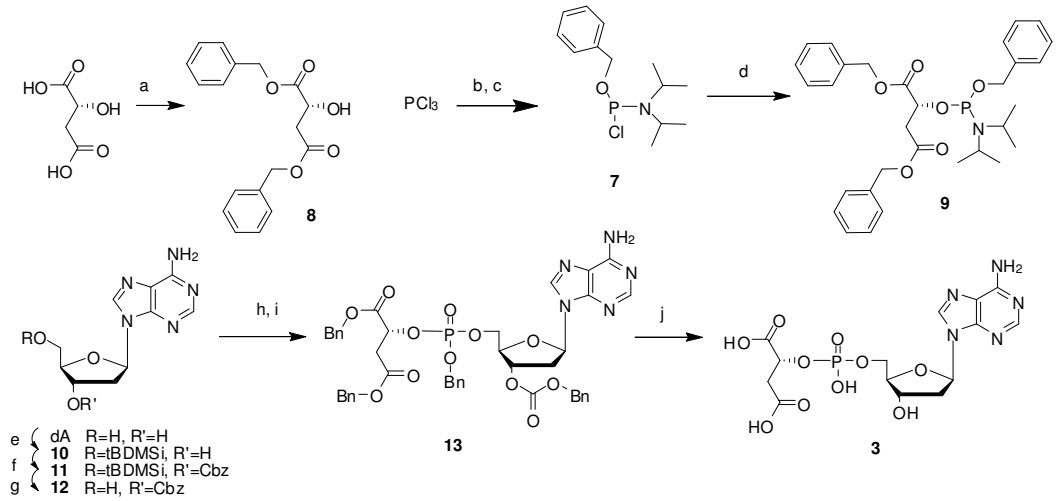



**Scheme III.1:** Synthesis of malic acid dAMP *via* benzyl protected phosphoramidite approach: a) benzyloxydiisopropylamino chlorophosphite, DIPEA, anh. acetonitrile; b) tetrazole, dimethyl ester L-malic acid ; acetonitrile anh.; c) iodine solution in pyridine, THF, water ; d) potassium bicarbonate, methanol, water ; e) palladium on carbon, activated, methanol.

Mild basic treatment of **6** to remove of the carboxylate protecting methyl esters resulted in complete hydrolysis of the compound. As a consequence, we chose to protect the dicarboxylic acid with benzyl moieties, allowing a milder final deprotection reaction. Unfortunately, confronted to the bulky dibenzylated malic acid, **5** proved unable to serve as electrophilic centre, probably due to steric hindrance. The more reactive diethylamino phosphorous chloride or the stronger activation reagent, pyridinium trifluoroacetate were employed to enhance the electrophilicity of the phosphorous in the first or the second step, respectively; no evidence of coupling was found. The steric hindrance of the phosphate protecting benzyl group was probably a reason for this lack of success, since cyanoethyl protected phosphoramidite chemistry was successful in yielding the phosphotriester (results not shown).

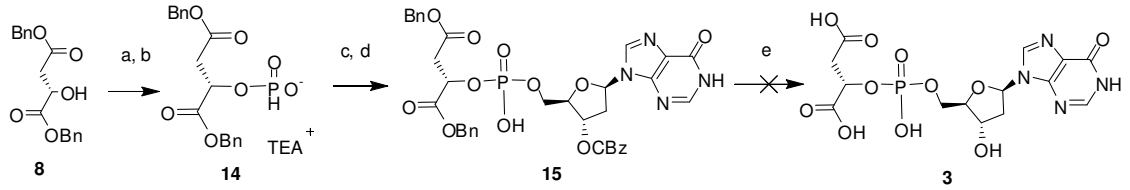
Hence, we decided to prepare the phosphoramidite reagent from benzylated malic acid and react it with the more nucleophilic 5′-hydroxyl of deoxyadenosine, as shown in scheme III.2. The benzyloxycarbonyl group (Cbz), cleavable by hydrogenation, was chosen to protect the 3′-hydroxyl group (99). Transient *tert*-butylsilyl protection of the 5′-OH of deoxyadenosine

(compound **10**) followed by benzyloxycarbonylation of the 3′-OH (compound **11**) and final fluoride deprotection of the 5′-OH was adapted from published procedures (99), affording 3′-O-Cbz-2′-deoxyadenosine **12**. Hereto, malic acid was benzylated under acidic conditions (compound **8**). The phosphorous chloride specie **7** was prepared in a one-pot synthesis, during which substitution of one chloride by benzyl alcohol group was followed by the substitution of the second chloride with a diisopropylamino group. Key step phosphorous coupling to **8** was done in dry acetonitrile in the presence of diisopropylethylamine (DIPEA) to afford phosphoramidate **9**. $^{31}$P NMR showed the expected doublet signal around 150 ppm, indicating the presence of a phosphoramidite species. Subsequent nucleophilic attack of the 5' hydroxyl group of 3′-Cbz-dA afforded the desired phosphittriester. Standard iodine oxidation in presence of pyridine only led to isolation of the benzyl protected monophosphate of Cbz-dA, suggesting that the phosphittriester or the phosphotriester species was formed and further degraded. *Tert*-butylperoxide oxidation carried out in a mixture of decane and dichloromethane as developed by Hayakawa (21) could prevent this rapid degradation and afforded the desired phosphotriester **13** in small amounts. $^{31}$P NMR analysis showed a typical diastereoisomeric mixture profile (doublet with approx. -2 ppm shift). Hydrogenation in presence of palladium on carbon afforded the desired phosphodiester **3**. Unfortunately the time spent in solution for characterization purposes led to phosphodiester cleavage.



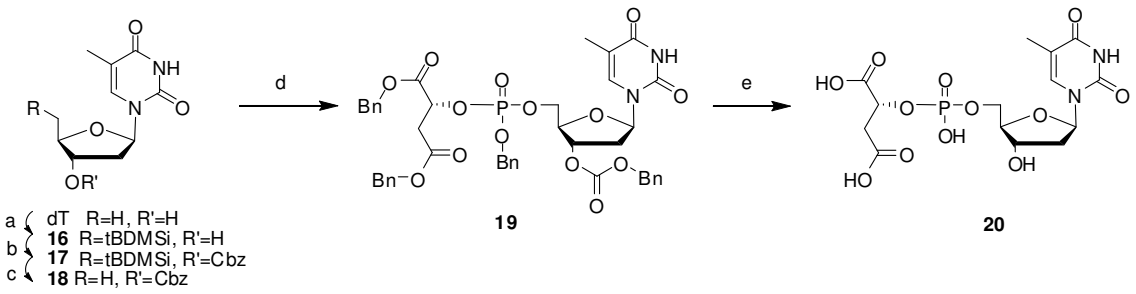

**Scheme III.2:** Synthesis of **3** *via* phosphite triester approach: a) Benzyl alcohol, p-TsOH, toluene, reflux; b) DIPA, diethylether, anhydrous, 0°C; c) benzyl alcohol, anhydrous; d) DIPEA, **8**, acetronitrile, anhydrous, 0°C; e) *tert*-butyldimethylsilyl chloride, pyridine anh., rt; f) benzylchloroformate, pyridine, anh g) triethylammonium hydrogen fluoride, N-methyl pyrrolidone, triethylamine, rt; h) tetrazole **8**, dichloromethane, distilled, rt; i) ter-BuOOH, decane, dichloromethane, 0°C ; j) hydrogen, palladium on carbon, activated, ethanol, rt.

Simultaneously and in the hope for a more successful coupling between benzylated malic acid phosphate and deoxyadenosine, the H-phosphonate coupling strategy was explored and is described in scheme III.3. The H-phosphonate of alcohol **8**, compound **14**, was obtained *via* pivaloyl chloride activation in the presence of pyridine, adapting the approach developed by the Stawinski group (100). An additional equivalent of pivaloyl chloride promoted the coupling to deoxyadenosine (101). Presence of the benzyl and benzyloxycarbonyl protected compound **15** was witnessed by $^{31}P$ and mass spectrometry. Unfortunately, numerous precautions in order to prevent product degradation were unsuccessful before or after hydrogenation of the benzyl groups.



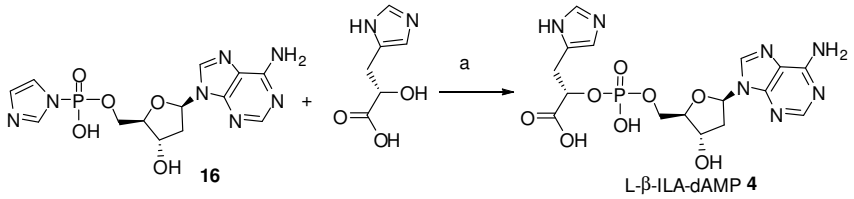

**Scheme III.3:** H-phosphonate chemistry starting with dibenzylated malic acid: a) phosphonic acid, pivaloyl chloride, pyridine, anhydrous, 0°C; b) triethylammonium carbonate solution 1M, r.t.; c) **12**, pivaloyl chloride, pyridine, anhydrous, 0°C; d) *tert*-butyl peroxide in decane 6.0 M, 0°C; e) hydrogen, $P_{atm}$, palladium on carbon, activated, ethanol, r.t.. Bn = benzyl Cbz = benzyloxycarbonyl.

Stability studies carried out on amino acid dAMP phosphoramidates by Dr. Dyubankova in the laboratory for medicinal chemistry showed that thymidine analogues presented with higher half life in solution (Herdewijn, unpublished results). The thymidine analogue malic acid dTMP (**20**) was therefore synthesized resorting to an identical H-phosphonate synthetic route (Scheme III.4). Rapid degradation of **20** (towards the benzyl phosphodiester of dTMP) occurred during and/or after the final hydrogenolysis step. In summary, the phosphodiester linkage involving malic acid proved to undergo hydrolysis too rapidly and could therefore not be used for *in vitro* testing.




**Scheme III.4:** Synthesis of the thymidine analogue *via* phosphite triester approach : a) *tert*butyldimethylsilyl chloride, pyridine anhydrous ; b) benzylchloroformate, pyridine, anhydrous c) triethylammonium hydrogen fluoride, THF ; d) tetrazole, **8**, dichloromethane distilled; e) hydrogen, palladium on carbon activated, ethanol.

Since L-β-imidazole lactic acid-dAMP **4** features an imidazole group in place of a carboxylic acid function in β position of the linking atom, it was reasoned that it could be potentially more stable than L-malic acid dAMP **3** towards hydrolysis. The divalent cation assisted coupling proposed by Sawai was used to synthesize the new candidate (Scheme III.5) (85). Starting reagent dAMP imidazolate **16** was prepared as described in chapter II. Due to the additional formation of a diphosphate species and simultaneous solvolysis, the desired product could be obtained only in moderate yield (50%) after ion exchange chromatography.




**Scheme III.5**: Synthesis of L-β-imidazole-lactic acid dAMP: a) zinc chloride, N-ethyl morpholine buffer pH 7.5.

## III.2.3 Single incorporation experiments

As for the series of analogues presented in chapter two, we evaluated **2** and **4** as substrates of HIV-1 reverse transcriptase for nucleotide incorporation into a dsDNA duplex. Previous experiments carried out with the L-aspartic acid and L-histidine phosphoramidates of dAMP demonstrated that these amino acids were acceptable leaving groups for the polymerase in the nucleotidyl insertion process.

Here we evaluated the capacity of HIV-1 RT to recognize and incorporate a deoxyadenosine nucleotide into the primer-template duplex P1:T1 (Table 1) when the pyrophosphate group is replaced by L-β-imidazole-lactic acid as compared to L-His- and natural PPi. Incorporation efficiency was analyzed by the polyacrylamide gel-based single nucleotide incorporation assay (87;88).

**Table III.1.** Overview of the primer-template duplexes used in the DNA polymerase reactions. Bold letters indicate the template overhang in the hybridized primer-template duplex.

| | |
|---|---|
| Single nucleotide incorporation and kinetic experiments | |
| P1 | 5'-CAGGAAACAGCTATGAC-3' |
| T1 | 3'-GTCCTTTGTCGATACTG**TCCC**-5' |
| Elongation experiments | |
| T2 | 3'-GTCCTTTGTCGATACTG**TTTTTTTGGAC**-5' |

Replacement of the nitrogen (in L-His-dAMP **2**) by an oxygen atom (in L-ILA-dAMP **4**) resulted in improved single incorporation. Using L-ILA-dAMP as a substrate we observed up to 60% of extended primer at 0.5 mM, whereas with L-His-dAMP the incorporation only yielded 50% of extended primer (Figure III.3). Surprisingly, for **4** we observed better incorporation at 5- and 10-fold lower concentrations (at a 100 µM deoxynucleotide concentration, conversion of a primer to a (P+1) strand ranged from 80% to completion in 20 min (Figure III.4) (102).


**Figure III.3:** Profile of incorporation of dAMP into P1:T1 by HIV-1 reverse transcriptase using L-ILA-dAMP **4** (500 µM), L-His-dAMP **2** (500 µM) and dATP (10 µM) as substrates. Aliquots were drawn between 10 and 120 minutes.

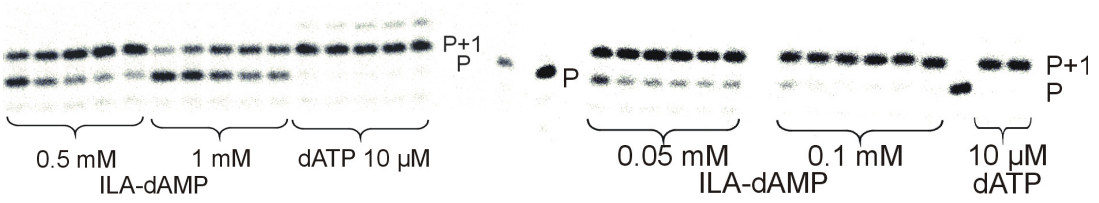
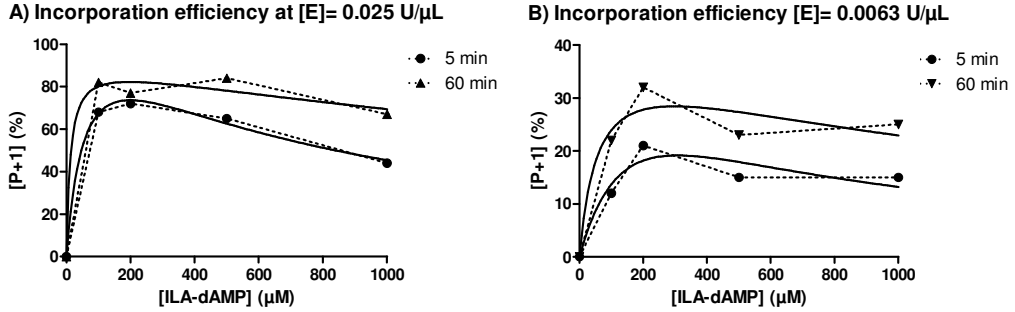



**Figure III.4**: Profile of incorporation of dAMP into P1:T1 by HIV-1 reverse transcriptase using ILA-dAMP **4** as substrate at 500 and 1000 µM (left panel) and at 50 and 100 µM (right panel) concentrations. Aliquots were drawn between 10 and 60 minutes.

These results suggested that the reaction could be inhibited by the substrate L-ILA-dAMP. Incorporation efficiency was plotted in function of substrate concentration for two different enzyme concentrations (0.025 U/µL, graphic III.1A and 0.0063 U/µL, III.B). All profiles showed a decrease in incorporation efficiency for substrate concentrations above 200 µM, supporting a substrate inhibition eventuality. An accurate determination of inhibition parameters could be done to confirm this hypothesis by plotting $K_M$ in function of the substrate concentration, but such fundamental investigation is not the purpose of our study.




**Graphic III.1:** Incorporation efficiency (as % of elongated primer) measured at 5 and 60 min reaction time in function of substrate concentration for **4** A) 0.025 U/µL and B) 0.0063 U/µL HIV-1 RT (Ambion, specific activity 11.700 U/mg).

Therefore, we compared both substrates at lower concentrations, to fall out of the range of possible substrate inhibition concentrations (approximately 0.25 mM). At 125 and 250 µM (Figure III.5), incorporation with **4** was more efficient (40% (P+1)) than with L-His-dAMP (17% (P+1)).

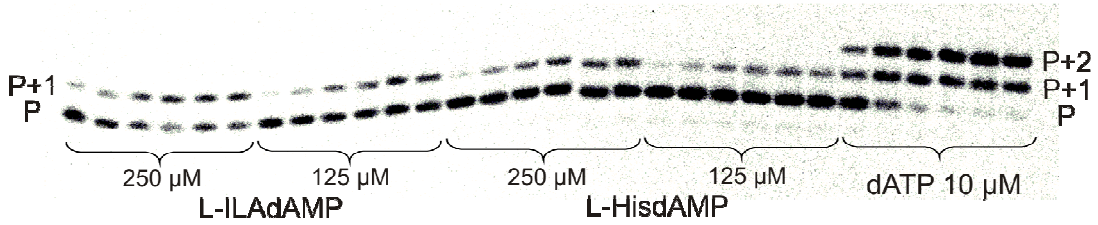



**Figure III.5:** (From l. to r.) incorporation of L-ILA-dAMP **4** (250 and 125 µM), L-His-dAMP **2** (250 and 125 µM) and dATP (10 µM) into P1T1 by HIV-1 reverse transcriptase. Aliquots were drawn between 10 and 120 minutes.

## III.2.4 Elongation experiments

Afterwards, we evaluated L-His-dAMP and L-ILA-dAMP as substrates for primer elongation using the primer-template duplex P1:T2 (Table III.1). Elongation efficiency was somewhat better with L-ILA-dAMP as substrate (Figure III.6). Nevertheless, complete strand elongation could not be obtained for either triphosphate mimic at 0.1 mM substrate concentration, nor could it at 0.5 mM substrate concentration (picture not shown).

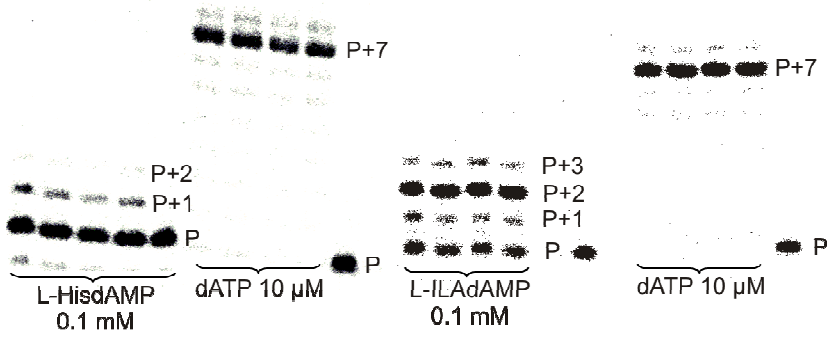



**Figure III.6:** Elongation of duplex P1:T2 with L-His-dAMP and L-ILA-dAMP at 100 µM concentration and reference compound dATP at 10 µM by HIV-1 RT (Ambion, specific activity 11.700 $U.mg^{-1}$). Aliquots were drawn at 30, 60, 90 and 120 minutes.

The premature termination of DNA synthesis can have many causes. Since stalling of HIV-1 RT is observed for both L-His- and L-ILA-dAMP, it cannot be attributed to the nature of the linkage. It is more likely due to the properties of the leaving group. Elongation tests involving a different leaving group, such as iminodiacetic acid, linked *via* a P-N bond to dAMP (IDA-dAMP, figure III.7. See corresponding study in chapter IV) resulted in a reduced stalling of the polymerase. With IDA-dAMP as substrate the full-length oligonucleotide could be

obtained as one of the major products, supporting a leaving group induced pausing hypothesis (see Chapter IV).

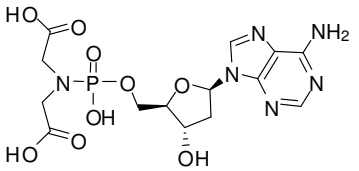


**Figure III.7:** Molecular structure of iminodiacetic acid dAMP

DNA synthesis termination could also be inherent to the nucleotide sequence. Several polymerases are known for pausing and strand termination when encountering hairpin structures (for example HIV-1 RT (103)), or sequences such as pyrimidine-GC motifs (T7 DNA polymerase (104)), repeated trinucleotides sequences (Pol I) (105;106) and poly(T)/poly(dA) sequences (HIV-1 RT (107)). The use of a poly(T) template for evaluation of poly(A) incorporation could therefore account for the decrease in elongation efficiency when using unusual substrates like L-ILA-, L-His- and L-Asp-dAMP. This hypothesis could be challenged by confronting this incorporation system with upstream and downstream differing dsDNA duplexes.

On the other hand, the polymerization reaction is generally thought to be driven forward by the hydrolysis of the PPi leaving group (3;93). Yin and Steitz suggested that the absence of the energetically rewarding hydrolysis of PPi might be responsible for less efficient and less selective nucleotide incorporation by T7 RNA polymerase (108). Using a non hydrolysable *β*-imidazole lactic acid instead of PPi could therefore be another cause for the decreased catalytic efficiency of HIV-1 RT. In other terms, the absence of released PPi and associated 'cellular energy' in the mechanism might account for the loss of processivity and/or efficiency of reverse transcriptase. In order to explore this lead, additional elongation experiments were performed. Temperature modulation was excluded since experimentation temperature is fixed as the optimal working temperature of the enzyme, in our case 37°C. Therefore, elongation reactions were performed in the presence an 'external' source of pyrophosphate (PPi, ATP or dATP) in various micromolar concentrations. Introduction of PPi or ATP at time zero or after 60 minutes reaction did not allow DNA synthesis to resume. If their presence promoted the reverse reaction (pyrophosphorolysis of the preceding 3' nucleotide), the corresponding kinetic equilibrium was too rapid to be visible with our imaging method. When the natural substrate of the enzyme dATP was introduced after 60 minutes reaction with the modified

substrate, chain elongation resumed, proving that reverse transcriptase retained the ability to elongate a primer strand in standard conditions.

### III.2.5 Kinetic experiments

Maximum velocity of the reaction (Vmax), affinity ($K_M$) and efficiency ($V_{Max}/K_M$) of the triphosphate analogues as substrates for HIV-1 RT were determined on the basis of the single completed hit model and are presented in Table III.2. P1 and T1 were used respectively as primer and template DNA strands (Table III.1).

**Table III.1.** Kinetic parameters of the incorporation of dAMP into P1T1 by HIV-1 RT DNA polymerase at 0.0063 U/µL (dATP and ILA-dAMP) or 0.025 U/µL (L-His-dAMP)

| | linkage | $K_M$ [µM] | $V_{max}$ [pmol.$U^{-1}$.$min^{-1}$] | $Vmax/K_M$ Substrate efficiency |
|---|---|---|---|---|
| dATP | ref | 0.48 ± 0.13 | 5.114 ± 0.397 | 10.65 |
| L-His-dAMP | P-N | 505.0 ± 114.8 | 0.330 ± 0.033 | 0.0007 |
| ILA-dAMP | P-O | 204.7 ± 30.8 | 2.746 ± 0.144 | 0.013 |

Both dNTP analogues were processed by HIV-1 RT and displayed a Michaelis-Menten kinetic profile. In the case of structurally identical side chains, the replacement of the linking nitrogen by an oxygen atom appeared to have two beneficial effects: substrate affinity was doubled and maximum velocity was increased by a factor 10. In other terms, during incorporation a phosphodiester linkage is being more efficiently cleaved by HIV-1 RT than a phosphoramidate linkage. This is consistent with the incorporation results obtained with the series of non-amino-acid-like candidates (Chapter II). Changing the nature of this linkage is therefore a way to modulate the kinetic parameters of the leaving group release reaction and can be taken in account for the design of other alternative leaving group candidates. However, implementation of this concept on pre-existing successful phosphoramidate analogues is likely to be limited by stability issues, as observed for the L-malic acid dAMP, phosphodiester analogue of L-aspartic acid dAMP.

### III.2.6 Exploration of other possible linkages

As mentioned in table I.2, other chemical entities such as sulphur (in C-S-P) and methylamine (in C-N($CH_3$)-P) are isosteric to NH in the linkage between phosphorous atom and carbon atom. The methyl amino analogue of L-Asp-dAMP (figure III.8) was previously

synthesized and tested for enzymatic incorporation by HIV-1 RT (83). Unfortunately, this deoxynucleoside triphosphate analogue was not processed by the enzyme in the usual test conditions.

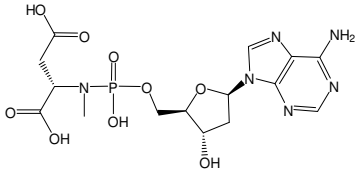


**Figure III.8:** Chemical structure of N-Methyl-(L-) aspartic acid dAMP phosphoramidate.

The sulphur analogues of β-Ala-dAMP (Chapter II, compound **4**), glycolic acid-dAMP (ibid., compound **5**) and L-Asp-dAMP are presented in figure III.9.

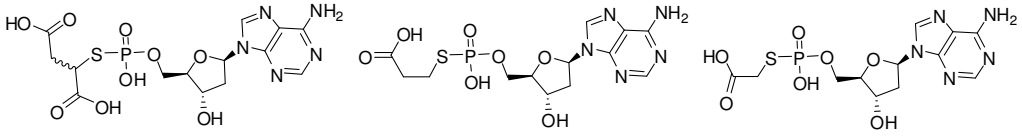


**Figure III.9:** Chemical structures of: (from left to right) DL-thiomalic acid, analogue of DL-Asp-dAMP; thiopropionic acid-dAMP (analogue of **4**) and thioacetic acid-dAMP (analogue of **5**).

Attempts to synthesize these dAMP phosphorothioates were made resorting to both DCC coupling and divalent cation assisted coupling procedures, with commercially available thiol compounds. With both procedures presence of the dimer dAppdA was immediately witnessed (Figure III.10), and no evidence of formation of the phosphorothioate analogues was found by the usual analytic methods. This absence could be explained by rapid dimerization of dAMP imidazolide, as already observed during the synthesis of GA-dAMP **5** (see scheme II.5). It is also possible that the phosphorothioate species is formed as an instable intermediate and functions as an activated dAMP. In other words, the thiolate groups seem to be to good leaving groups to allow isolation of the corresponding dAMP phosphorothioate analogues.

## *III.3 Discussion*

In order to evaluate the influence of the nature of the leaving group connecting atom on the kinetics of deoxyadenosine monophosphate incorporation by DNA polymerases, the

phosphoramidate L-His-dAMP and its phosphodiester counterpart L-ILA-dAMP were synthesized and evaluated as substrates for HIV-1 reverse transcriptase. Both dATP analogues were used by the enzyme to obtain 50-60 % (L-His-) to 80-100 % (L-ILA-) of an extended primer after 60 min reaction. In other terms both alternative structures were able to mimic the pyrophosphate moiety as leaving group in the polymerization reaction. The difference in incorporation efficiency between the analogues seems to be due to an increase in both the apparent affinity and the maximum velocity when the phosphoramidate linkage is replaced by a phosphodiester linkage. This is consistent with preliminary results obtained with various non-amino-acid-like leaving groups. Full-length elongation templated by seven complementary bases could not be obtained with either of the dATP analogues. Therefore, premature DNA synthesis termination cannot be attributed to the nature of the linkage between leaving group and deoxynucleoside. The influence of the leaving group's linkage on the kinetics of the incorporation reaction could be used to increase or slow down the rate of leaving group release and help understanding the chemical steps of DNA polymerization. Full understanding of the characteristics and influence of the leaving group in DNA polymerization could help designing new antiviral compounds, or be used in the context of nucleic acid synthesis orthogonalization. The instability of phosphorothioate analogues of both amino acid-dAMP did not allow the testing of a phosphorothioate linkage in this study.

## *III.4 Experimental section*

### III.3.1 General methods

NMR spectral analyses were carried out and processed as described in Chapter II. Chemical shifts are expressed in parts per million (*ppm*) by frequency. $^{1}H$ chemical shifts are referenced to an internal TMS signal ($\delta$ = 0.00 ppm), $^{31}P$ NMR chemical shifts are referenced to an external 85 % $H_3PO_4$ standard ($\delta$ = 0.00 ppm). Standard and exact mass spectra were measured as described in Chapter II. Chemicals of analytical or synthetic grade were obtained from commercial sources and were used as received unless otherwise specified (Dimethylaminopyridine, L-malic acid, para-toluene sulfonic acid sodium salt, *tert*-butanol, N-ethyl morpholine, triethylamine, N-ethylmorpholine, *tert*-butyl dimethyl chlorosilane, benzylchloroformate, N-methyl pyrrolidone: Acros; sodium hydroxide, phosphorous trichloride, DIPEA, diisopropylamine, benzyl alcohol, triethylammonium hydrogen fluoride, *tert*-butyl peroxide in decane 6.0 M, 10% wt. palladium on activated carbon Degussa type

E101 NE/W, phosphonic acid: Aldrich; L-histidine methyl ester hydrochloride, Alfa Aesar; imidazole, triphenylphosphine, zinc chloride, ethylene diamine tetraacetate disodium salt and dicyclohexylcarbodiimide: Fluka; pivaloyl chloride: Janssen Chimica; deoxyadenosine monophosphate and β-imidazole lactic acid: Sigma Aldrich; Celite® 545: VWR Int.). Technical solvents were obtained from Brenntag (Deerlijk, Belgium).

**DNA duplexes preparation and polyacrylamide gel-based electrophoresis** were carried out as described in Chapter II.

## III.3.2 Synthesis

***L-malic acid dibenzyl ester (8)***

To a suspension of L-malic acid (8g, 59.7 mmol) and benzyl alcohol (12 mL, 119.4 mmol, 2 eq.) in anhydrous toluene (distilled over sodium) was added paratoluene sulfonic acid sodium salt (1.159 g, 5.97 mmol, 0.1 eq.). The reaction mixture was heated at reflux (140°C) for 6 hours. Water was separated and collected. Solvents were evaporated and the mixture was dissolved in dichloromethane and washed with a saturated solution of sodium hydrogencarbonate. Organic phases were evaporated under reduced pressure, affording 8 in quantitative yield. Rf ($CHCl_3$/$CH_3OH$:6/1) = 0.82. $^1$H NMR (300 MHz, $CDCl_3$): $\delta$ 7.36 (m, 10H, Ph), 5.16 (2s, 2H, C*H2*-Ph), 5.11 (s, 2H, C*H2*-Ph), 3.89 (dd, 1H, C$H_\alpha$), 3.2 (d, 1H, O$H$), 2.85 (m, 2H, C$H_{2\beta}$). $^{13}$C NMR (125 MHz, $CDCl_3$): δ 174.02 (CO), 170.95 (CO), 135.56 ($C_p$-Ph), 128.62-128.24 (C-Ph), 67.11, 66.65 ($CH_2$-Ph), 51.34($CH_\alpha$), 38.98 ($CH_{2\beta}$). MS (ESI) m/z calculated for $C_{18}H_{18}O_5$ 314.33, found 314.0.

***Benzyl alcohol diisopropylamino chlorophosphite (7)***

Phosphoroustrichloride (20 mL, 0.23 mol) was mixed with dried acetonitrile (10 mL) in a two neck 50 mL flask under argon atmosphere, and dried benzyl alcohol (3.5 mL) was added over 10 min *via* syringe. The resulting mixture was stirred for 30 min at room temperature. Solvent and excess of $PCl_3$ was distilled off. $^1$H NMR (300 MHz, $CDCl_3$ mol. sieves 3Å): $\delta$ 7.40 (m, 5H, Ph), 5.29 (s, 1H, $CH_2$), 5.27 (s, 1H, $CH_2$) ppm. $^{13}$C NMR (125 MHZ, $CDCl_3$ mol. sieves 3 Å): $\delta$ 128.78-128.42 (Ph), 69.80 ($^2J_{C\text{-}P}$= 9.64 Hz, $CH_2$-Ph) ppm. $^{31}$P NMR (121 MHz, $CDCl_3$ mol. sieves 3 Å): $\delta$ 177.09 ppm.

Diethylether (10 mL) was added to the resulting white mixture under argon atmosphere and this solution was cooled down to 0°C and a solution of diisopropylamino (7.6 mL) in diethylether (10 mL) was added dropwise under vigorous stirring. The reaction mixture was let to warm up to r.t and stirred overnight. The white precipitate of diisopropylamonium chloride was filtered off under argon. Diethylether was then evaporated under reduced pressure at 40°C for 4 hours affording **7** in quantitative yield. $^{1}$H NMR (300 MHz, $CDCl_3$ mol. siev. 3Å): $\delta$ 7.36 (m, 5H, Ph), 4.92 (s, 1H, $CH_2$), 4.89 (s, 1H, $CH_2$), 3.84 (sept., 2H, C*H*-$(CH_3)_3$), 1.27 (m, 12H, CH-(C*H*$_3$)$_3$) ppm. $^{13}$C NMR (125 MHZ, $CDCl_3$ mol. siev. 3 Å): $\delta$ 128.86-127.27 (Ph), 67.30 ($^{2}J_{C\text{-}P}$ =18.18 Hz, P-*C*$H_2$), 46.12 ($^{1}J_{C\text{-}P}$ =12.95 Hz, P-*C*H) 23.62 (CH-(*C*$H_3$)$_3$) ppm. $^{31}$P NMR (121 MHz, $CDCl_3$ mol. siev. 3 Å): $\delta$ 181.53 ppm.

***(L-malic acid dibenzyl ester) benzyl diisopropylamino phosphoramidite (9)***

A solution of 7 (192 mg, 0.7 mmol) in anhydrous acetonitrile (5 mL) was cooled down to -10°C under argon atmosphere. Upon cooling, diisopropylethylamine (236 µL, 1.5 mmol, 2 eq.) was added *via* syringe, followed by 8. The resulting clear mixture was let to warm up to room temperature for 1.5 hour then solvents were evaporated under reduced pressure. The resulting oil was dissolved in ethyl acetate and washed with water. Organic phases were gathered and evaporated, and the residue was subjected to flash silica chromatography eluting a gradient of ethyl acetate in hexane containing 0.1 % triethylamine. Corresponding fractions were collected and evaporated, yielding 110 mg of colourless oil (0.20 mmol, 20% yield) which was stored in anhydrous dichloromethane and under argon. $^{1}$H NMR (300 MHz , $CDCl_3$): $\delta$ 7.33-7.30 (m, 15H, Ph), 5.14 (s, 2H, C*H*$_2$-Ph), 5.10 (s, 2H, C*H*$_2$-Ph), 5.13 (s, 2H, C*H*$_2$-Ph), 5.06 (s, 2H, C(O)-C*H*$_2$-Ph), 4.61 (m, 1H, C*H*$_\alpha$), 3.63 (sept, 2H, C*H*-$(CH_3)$, 22.91 (m, 2H, C*H*$_{2\beta}$), 1.16 (2s, 12H, CH-(C*H*$_3$)$_2$. $^{31}$P NMR (121 MHz, $CDCl_3$): $\delta$ 150.88, 150.69 (Rp and Sp). HRMS (positive ion mode): m/z calculated for $C_{31}H_{38}NO_6P$ 551.24367, found 552.2512.

***5'-tertbutylsilyl-2'-deoxyadenosine (10)***

2′-deoxyadenosine (2.0 g, 8.0 mmoles) was co-evaporated three times with freshly distilled pyridine to remove traces of water. It was redissolved in pyridine (20 mL) and *tert*-butylchlorosilane (1.32 g, 8.8 mmoles, 1.1 eq.) was added in one portion. The resulting mixture was stirred overnight at r.t. under argon atmosphere. Solvent was evaporated and coevaporated three times with toluene. The crude mixture was redissolved in chloroform, and extracted with water. The organic phase was collected and dried over sodium sulfate. After

filtration and evaporation, a white solid was obtained (2.56 g, 7 mmoles, 88 % yield). The crude was used as such for the next step. Rf($CHCl_3$/$CH_3OH$/$Et_3N$:6/1/0.1) = 0.5. $^1H$ NMR (300 MHz, $CDCl_3$): $\delta$ 8.35 (s, $1H_2$), 8.18 (s, 1H, $H_8$), 6.53 (apparent t, $^3J_{H1'-H2'}$ = 6.47 Hz, 1H, $H_{1'}$), 5.88 (bs, 2H, N$H_2$), 4.70 (m, 1H, $H_{3'}$), 4.12 (m, 1H, $H_{4'}$), 3.89 (m, 2H, $H_{5'}$), 2.71 (m, 1H, $H_{2'}$), 2.60 (m, 1H, $H_{2''}$), 0.92 (s, 9H, C-(C$H_3$)$_3$), 0.11 (s, 6H, Si-C$H_3$) ppm. $^{13}C$ NMR (125 MHz, $CDCl_3$): $\delta$ 155.4 ($C_6$), 153.0 ($C_2$), 149.5 ($C_4$), 138.9 ($C_8$), 119.9 ($C_5$), 87.1 ($C_{4'}$), 84.1 ($C_{1'}$), 72.1 ($C_{3'}$), 63.4 ($C_{5'}$), 41.3 ($C_{2'}$), 26.0 (C-(*C*$H_3$)$_3$), 18.4 (Si-*C*-($CH_3$)$_3$), -5.40 (Si-*C*$H_3$) ppm.

***3'-benzyloxycarbonyl-5'-tertbutylsilyl-2'-deoxyadenosine (11)***

After **10** (1.46 g, 4 mmoles) was dissolved in freshly distilled dichloromethane (25 mL), benzylchloroformate (0.85 mL, 6 mmoles, 1.5 eq.) and dimethylaminopyridine (1.5 g, 12 mmoles, 3 eq.) and the resulting mixture was stirred at r.t. for 4 days. The mixture was diluted with dichloromethane and extracted with 2M hydrogen chloride solution. The organic phase was collected, dried with sodium sulfate. Finally, the solvent was evaporated after filtration. The crude mixture was purified by silica chromatography using a gradient of methanol (0 to 2.5 %) in dichloromethane. **11** was obtained as a white solid (1.43 g, 2.8 mmoles, 72% yield). TLC $CHCl_3$/$CH_3OH$ 0.75. $^1H$ NMR (300 MHz, $CDCl_3$): $\delta$ 8.37 (s, $1H_2$), 8.20 (s, 1H, $H_8$),7.40 (bs, $H_{aro}$ Cbz), 6.52 (apparent t, $^3J_{H1'-H2'}$ = 6.68 Hz, 1H, $H_{1'}$), 6.02 (bs, 2H, N$H_2$), 5.36 (m, 1H, $H_{3'}$), 5.21 (s, 2H, $CH_2$ Cbz), 4.31 (m, 1H, $H_{4'}$), 3.93 (m, 2H, $H_{5'}$), 2.73 (m, 2H, $H_{2'}$), 0.92 (s, 9H, C-(C$H_3$)$_3$), 0.11 (s, 6H, Si-C$H_3$) ppm. $^{13}C$ NMR (125 MHz, $CDCl_3$): $\delta$ 155.30 ($C_6$), 154.41 (Ph), 153.00 ($C_2$), 149.74 ($C_4$), 138.59 ($C_8$), 134.79 (Ph), 128.71-128.38 (Ph), 120.30 ($C_5$), 85.24 (C4'), 84.17 ($C_{1'}$), 78.93 ($C_{3'}$), 70.10 ($CH_2$), 63.57 (C5'), 38.95 ($C_{2'}$), 25.93 (C-(*C*$H_3$)$_3$), 18.36 (Si-*C*-($CH_3$)$_3$), -5.48 (Si-*C*$H_3$) ppm.

***3'- benzyloxycarbonyl -2'-deoxyadenosine (12)***

The product **11** (400 mg, 0.8 mmoles) was dissolved in N-methylpyrolidone (2 mL) under argon atmosphere and triethylamine (1mL), then triethylammonium hydrogen fluoride (1.34 mL) were added *via* syringe to obtain a clear yellow solution. After stirring at r.t. fro 30 min, the mixture was diluted with 50 mL ethyl acetate, washed twice with 25 mL water. The organic phases were gathered, dried with $Na_2SO_4$ and evaporated. Product **12** was obtained as a white powder (520 mg, 1.35 mmol, 67% yield). TLC ($CHCl_3$:$CH_3OH$/6:1) Rf=0.53.

$^{1}$H NMR (300 MHz, DMSO $d_6$): $\delta$ 8.37 (s, 1H, $H_2$), 8.17 (s, 1H, $H_8$),7.40 (bs, $H_{aro}$ Cbz), 7.54 (bs, 2H, N$H_2$), 6.36 (apparent t, $^{3}J_{H1'-H2'}$ = 5.87 Hz, 1H, $H_{1'}$), 5.34 (m, 1H, $H_{3'}$), 5.21 (s, 2H, $CH_2$ Cbz), 4.16 (bs, 1H, $H_{4'}$), 3.64 (m, 2H, $H_{5'}$), 3.01 (m, 1H, $H_{2'}$), 2.73 (m, 2H, $H_{2'}$) ppm. $^{13}$C NMR (125 MHz, DMSO-$d_6$): $\delta$ 156.06 ($C_6$), 154.25 (Ph), 152.18 ($C_2$), 149.22 ($C_4$), 140.37 ($C_8$), 135.75 (Ph), 129.00-128.74 (Ph), 119.75 ($C_5$), 85.58 ($C_{4'}$), 84.59 ($C_{1'}$), 79.54 ($C_{3'}$), 69.72 ($CH_2$-Ph), 62.17 ($C_{5''}$), 38.88 ($C_{2'}$) ppm. MS (ESI positive ion mode): m/z calculated for $C_{18}H_{19}N_2O_5$ 385.37, found 386.1.

***3'-benzyloxycarbonyl-2'-deoxyadenosine-5'-(malic acid dibenzyl ester)-benzyl phosphotriester (13)***

To a solution of **12** (77 mg, 0.2 mmol, 1 eq) in anhydrous dichloromethane (2 mL) kept under argon atmosphere was added a solution of **9** (165 mg, 0.3 mmol, 1.5 eq) in anhydrous dichloromethane (1 mL). A solution of tetrazole in acetonitrile (0.1 M, 4 eq.) was added *via* syringe and the resulting mixture was stirred for 2 hours at room temperature. After completion of the reaction was assessed by TLC (hexane/acetone/TEA: 50/50/1), an oxidizing solution of *tert*-butanol peroxide in decane (6.0 M, 0.3 mL, 1 eq.) was added *via* syringe and the mixture was further stirred for 20 min. The mixture was then partitioned between dichloromethane and a saturated sodium chloride solution. Organic phase were gathered, evaporated and co-evaporated with dichloromethane 3 times. The yellow oily residue was purified by flash chromatography eluting with a gradient of methanol in dichloromethane and affording **13** in small amount (10 mg, 0.012 mmol, 5% yield). $^{1}$H NMR (300 MHz , $CDCl_3$): $\delta$ 8.34 (s, 1H, $H_2$), 8.13 (s, 1H, $H_8$), 7.32-7.28 (m, 20H, Ph), 6.46 (dd, $^{3}J_{H1'-H2'}$ = 5.64 & 8.73 Hz, 1H, $H_{1'}$), 5.63 (bs, 2H, N$H_2$), 5.20-5.0 (m, 10H, 4xC$H_2$-Ph, C$H_\alpha$, $H_{3'}$), 4.32 (m, 2H, $H_{5'}$), 4.25 (m, 1H, $H_{4'}$), 2.93 (m, 2H, C$H_{2\beta}$), 2.73 (m, 1H, $H_{2'}$), 2.60 (m, 1H, $H_{2'}$) ppm. $^{13}$C NMR (125 MHz, $CDCl_3$): $\delta$ 168.7 (CO malic acid), 155.4 ($C_6$), 154.2 (CO Cbz) 153.1 ($C_2$), 149.8 ($C_4$), 138.7 ($C_8$), 135.4-135.2 (Ph), 119.8 ($C_5$), 84.1 ($C_{1''}$, 82.8 ($C_4$), 78.5 ($C_3$), 72.1 ($C_5$ or $CH_2$ malic acid), 70.1 ($C_5$ or $CH_2$ malic acid), 67.9, 67.1, 67.0 (C$H_2$Ph), 37.7 ($C_2$) ppm. $^{31}$P NMR (121 MHz, $CDCl_3$): $\delta$ -1.97, -2.11 ppm (Rp and Sp). HRMS (ESI, positive ionization mode): m/z calculated for $C_{43}H_{42}N_2O_{12}P^+$: 852.2682, found: 852.2647.

***2'-deoxyadenosine-5'-(malic acid)-phosphodiester* (3)**

Phosphotriester **13** (100 mg, 0.12 mmol) was dissolved in absolute ethanol or ethyl acetate (2 mL), palladium on carbon 10% (Degussa® type, 38.3 mg, 0.036 mmol, 0.3 eq) was added and

the resulting suspension was stirred for 24 hours under hydrogen atmosphere (atmospheric pressure) while monitoring by TLC ($CHCl_3$/$CH_3OH$/$H_2O$:5/3/0.5). Upon completion (TLC shows quantitative yield), the suspension was filtered on Celite® 545, and washed several times with ethanol. This solution was lyophilized, affording **3** as a white residue. $^{31}P$ NMR (121 MHz): $\delta$ (major peak) -1.15 ppm. MS (ESI negative mode): m/z calculated for $C_{14}H_{18}N_5O_{10}P$ 447.07, found (major peak) 446.2. Stability in $D_2O$ is inferior to 15 minutes.

***Synthesis of (L-malic acid dibenzyl ester) H-phosphonate TEA salt (14)***

Compound **8** (250 mg, 0.8 mmol) and phosphorous acid (655 mg, 8.0 mmol, 8 eq) were co-evaporated 2 times with anhydrous pyridine, separately. **8** was redissolved in anhydrous pyridine (10 mL) and added to previously lyophilized phosphorous acid. After cooling down the mixture to 0°C, pivaloyl chloride (0.5 mL, 4.0 mmol, 5 eq) was added *via* a glass syringe. The reaction was let to warm up to r.t. After overnight stirring, the mixture was quenched with TEAB 2M, then was partitioned solvents between dichloromethane and TEAB 0.5 M. Organic phases were gathered, dried on sodium sulfate, evaporated and finally co-evaporated with toluene to yield **14** as a triethylamonium salt (366 mg, 96 % yield). Rf ($CHCl_3$/$CH_3OH$:6:1) = 0.23. $^1H$ NMR (300 MHz, DMSO-*d6*): $\delta$ 7.68 (s, 0.5 H, $^1J_{P\text{-}H}$ = 608.54 Hz, H-P), 7.32 (m, 10H, Ph), 5.65 (s, 0.5H, $^1J_{P\text{-}H}$= 608.54 Hz, H-P), 5.07 (s, 2H, C$H_2$-Ph), 5.13 (s, 2H, C$H_2$-Ph), 4.88 (dd, 1H, C$H_\alpha$), 2.99 (q, 6H, C$H_2$($Et_3NH^+$)), 2.83 (1H, 2H, C$H_{2\beta}$), 1.12 (t, 9H, C$H_3$($Et_3NH^+$)) ppm. $^{13}C$ NMR (125 MHz, DMSO-*d6*): $\delta$ 168.7 (C(O) malic acid), 155.4 ($C_6$), 170.95 ($C(O)_\alpha$), 169.80, ($C(O)_\beta$), 136.30 (Ph), 128.85-128.27 (Ph), 69.09 ($CH_\alpha$), 45.82 ($CH_2$($Et_3NH^+$)), 27.47 ($CH_3$Ph), 8.91 (C$H_3$($Et_3NH^+$)) ppm. $^{31}P$ NMR (121 MHz, DMSO-*d6*): $\delta$ 0.64 ppm. MS (ESI, negative ion mode): m/z calculated for C18H19O7P 378.31, found 377.0.

***3'-benzyloxycarbonyl-2'-deoxyadenosine-5'-(malic acid dibenzyl ester) phosphodiester triethylammonium salt (15)***

Compound **14** (87 mg, 0.18 mmol) was dissolved in anhydrous pyridine (1 mL). After cooling down the mixture to 0°C, pivaloyl chloride (25 µL, 0.2 mol, 1.1 eq) was added dropwise and the resulting mixture was stirred for 20 minutes. A solution of **12** (55 mg, 0.2 mmol, 1.1 eq) in pyridine (2 mL) was added dropwise. After stirring at r.t. for 2 hours TLC showed completion of the reaction. *Tert*-butyl peroxide (1 eq.) in decane or dichloromethane was added and the mixture was let to react for 20 minutes before evaporating all solvents. The

resulting oil was dissolved in dichloromethane and washed with water. Organic phases were gathered, dried on sodium sulfate and evaporated. The crude was subjected to silica column chromatography eluting with a gradient of methanol in chloroform containing 0.1 % triethylamine. Product fractions were collected and evaporated, affording **15** as a white solid (32 mg, 22%, 29% yield). Rf ($CHCl_3/CH_3Cl/H_2O$: 5:3:0.5)=0.76. $^1$H NMR (300 MHz , $CDCl_3$): $\delta$ 8.61 (s, 1H, $H_2$), 8.31 (s, 1H, $H_8$), 7.32 (m, 5H, $Ph_{Cbz}$), 7.30 (m, 10H, $Ph_{Bn}$), 6.57 (dd, $^3J_{H1'-H2'}$= 5.8 & 8.9 Hz, 1H, $H_{1'}$), 6.36 (bs, 2H, N$H_2$), 5.42 (d, H, $H_{3'}$), 5.19 (s, 2H, C$H_2$Ph), 5.18 (m, 1H, $CH_\alpha$), 5.06 (s, 2H, C$H_2$Ph), 5.06 (s, 2H, C$H_{2Cbz}$), 4.32 (m, 1H, $H_{4'}$), 4.15 (m, 2H, $H_{5'}$), 2.98 (m, 7H, C$H_2$($Et_3NH^+$) & C$H_{2\beta}$), 2.84 (m, 1H, $H_{2'}$), 2.60 (m, 1H, $H_{2'}$), 1.23 (t, 9H, C$H_3$($Et_3N$) ppm. $^{31}$P NMR (121 MHz, $CDCl_3$): $\delta$ -1.68 ppm. MS (ESI, negative ion mode): m/z calculated for $C_{36}H_{36}N_5O_{12}P$: 761.7, found: 760.4.

***2'-deoxyadenosine-5'-(malic acid) phosphodiester (3)***

Phosphodiester **15** (30 mg, 0.04 mmol) was dissolved in absolute ethanol (2 mL), palladium on activated carbon 10% (Degussa® type, 5 mg, 0.012 mmol, 0.3 eq.) was added and the resulting suspension was stirred for 2 days under hydrogen atmosphere (atmospheric pressure) while monitoring by TLC. Upon completion (TLC indicates quantitative yield) the suspension was filtrated on Celite® 545, and washed several times with ethanol. The solvent was evaporated, affording **3** as a white residue. Rf ($CHCl_3/CH_3Cl/H_2O$: 5:3:0.5)=0.51. $^{31}$P NMR (121 MHz, $D_2O$) of the crude mixture showed multiple signals around $\delta$ 0 ppm (degradation products) and a peak at $\delta$ -1.14 ppm, which had disappeared within the 30 min of the NMR experiment.

***5'-tertbutylsilyl-2'-deoxythymidine (17)***

2′-deoxythymidine (2.0 g, 8.2 mmoles) was co-evaporated three times with freshly distilled pyridine to remove traces of water. It was redissolved in pyridine (40 mL) and *tert*-butylchlorosilane (1.36 g, 9.02 mmoles, 1.1 eq.) was added in one portion. The resulting clear mixture was stirred for 2 days at r.t. under argon atmosphere. Solvent was evaporated and the mixture co-evaporated three times with toluene. The crude mixture was redissolved in chloroform and extracted with water and the organic phases were collected, co-evaporated with toluene then dichloromethane and dried over sodium sulfate. **17** was obtained as a white solid (3.06 g, 8.2 mmoles, quantitative yield). The crude was used as such for the next step. Rf($CHCl_3/CH_3OH/Et_3N$:6/1/0.1) = 0.68. $^1$H NMR (300 MHz, $CDCl_3$): $\delta$ 8.38 (s, NH-T), 7.26

(s, 1H, $^4J$H$_{5'}$-H$_{6'}$=1.24Hz, H$_6$), 6.39 (dd, $^3J_{H1'-H2'}$ = 6.93 Hz, 1H, H$_{1'}$), 4.49 (m, 1H, H$_{3'}$), 4.06 (m, 1H, H$_{4'}$), 3.88 (dd, 2H, H$_{5'}$), 2.13 (m, 2H, H$_{2'}$), 1.94 (s, 1H, $^4J$H$_{5'}$-H$_{6'}$= 1.17 Hz, CH$_3$), 0.95 (s, 9H, C-(C*H$_3$*)$_3$), 0.14 (s, 6H, Si-C*H$_3$*) ppm. $^{13}$C NMR (125 MHz, CDCl$_3$): $\delta$ 163.5 (C$_4$), 150.16 (C$_2$), 138.0 (C$_6$), 110.9 (C$_5$), 87.03 (C$_{4'}$), 84.90 (C$_{1'}$), 72.60 (C$_{3'}$), 63.55 (C$_{5'}$), 41.13 (C$_{2'}$), 25.92 (C-(*C*H$_3$)$_3$), 18.36 (Si-*C*-(CH$_3$)$_3$), 12.54 (CH$_3$-T), -5.38 (Si-*C*H$_3$) ppm. HRMS (ESI, positive ion mode): calculated for $C_{16}H_{29}N_2O_5Si^+$ 357.18408, found 357.1870.

***3'-benzyloxycarbonyl-5'-tertbutylsilyl-2'-deoxythymidine (18)***

After **17** (2.92 g, 8.2 mmoles) was dissolved in freshly distilled dichloromethane (40 mL), benzylchloroformate (2.33 mL, 16.4 mmoles, 2.0 eq.) and dimethylaminopyridine (4.01 g, 32.8 mmoles, 4 eq.) and the resulting clear mixture was stirred at r.t. for 3 days. It was then diluted with dichloromethane (50 mL) and extracted with 2M hydrogen chloride solution. The organic phase was collected, dried with sodium sulfate, filtrated, and volatiles were evaporated. The crude mixture was purified by silica chromatography using a gradient of methanol (0 to 2.5 %) in dichloromethane. **18** was obtained as a thick yellow foam (3.01 g, 6.14 mmoles, 75% yield). TLC CHCl$_3$/CH$_3$OH 0.84. $^1$H NMR (300 MHz, CDCl$_3$): $\delta$ 8.33 (s, 1 NH), 7.54 (s, 1H, $^4J$H$_{5'}$-H$_{6'}$=1.20 H$_6$),7.40 (bs, H$_{aro}$ Cbz), 6.37 (dd, $^3J_{H1'-H2'}$ = 7.44 Hz, 1H, H$_{1'}$), 5.20 (s, 3H, H$_{3'}$, C*H$_2$*-Ph), 4.21 (bs, 1H, H$_{4'}$), 3.93 (m, 2H, H$_{5'}$), 2.53 (m, 1H, H$_{2'}$), 2.15 (m, 1H, H$_{2'}$), 1.94 (s, 1H, $^4J$H$_{5'}$-H$_{6'}$= 1.14, CH$_3$), 0.95 (s, 9H, C-(C*H$_3$*)$_3$), 0.15 (s, 6H, Si-C*H$_3$*) ppm. $^{13}$C NMR (125 MHz, CDCl$_3$): $\delta$ 163.39 (C$_4$), 155.48 (*C*(O)-Cbz), 150.04 (C$_2$), 134.98 (*C*-C$_{Ph}$), 128.72-128.128.38 (C-*C$_{Ph}$*), 111.11 (C$_5$), 85.01 (C$_{4'}$), 84.68 (C$_{1'}$), 78.85 (C$_{3'}$), 70.07 (CH$_2$), 63.64 (C$_{5'}$), 38.12 (C$_{2'}$), 25.91 (C-(*C*H$_3$)$_3$), 18.32 (Si-*C*-(CH$_3$)$_3$), -5.46 (Si-*C*H$_3$) ppm. HRMS (ESI, positive ion mode): calculated for $C_{24}H_{35}N_2O_7Si^+$ 491.2208, found 491.2224.

***3'- benzyloxycarbonyl -2'-deoxythymidine (19)***

Product **18** (2.99 g, 6.1 mmoles) was dissolved in tetrahydrofurane (25 mL) under argon atmosphere and the mixture was cooled down to 0°C. Triethylammonium hydrogen fluoride (3.93 mg, 24.4 mmoles, 4 eq.) was added *via* syringe to obtain a clear yellow solution. After stirring at r.t. for 24 h, the mixture was diluted with 50 mL ethyl acetate then washed twice with a saturated solution of sodium hydrogenocarbonate. The organic phases were gathered, dried with $Na_2SO_4$ and evaporated to a yellow oil. Final co-evaporation afforded **19** as a white powder (2.15 g, 5.7 mmol, 94% yield). TLC (CHCl$_3$:CH$_3$OH/6:1) Rf=0.63. $^1$H NMR (300 MHz, DMSO d$_6$): $\delta$ 7.72 (s, 1H, $^4J$H$_{5'}$-H$_{6'}$=1.14Hz, H$_6$),7.40 (bs, H$_{aro}$ Cbz), 6.16 (apparent t,

$^{3}J_{H1'-H2'}$ =7.293Hz, 1H, $H_{1'}$), 5.23 (s, 3H, $H_{3'}$), 5.17 (bs, 3H, C$H_2$-Ph and NH), 4.04 (s, 1H, $H_{4'}$), 3.63 (m, 2H, $H_{5'}$), 2.34-2.31 (m, 2H, $H_{2'}$), 1.78 (s, 1H, $^{4}J$H$_{5'}$-H$_{6'}$= 0.87Hz,$CH_3$) ppm. $^{13}$C NMR (125 MHz, $CDCl_3$): $\delta$ 163.39 (C4), 155.48 (*C*(O)-Cbz), 150.04 (C2), 134.98 (*C*-$C_{Ph}$), 128.72-128.128.38 (C-$C_{Ph}$), 111.11 ($C_5$), 85.01 (C4′), 84.68 ($C_1$′), 78.85 ($C_3$′), 70.07 ($CH_2$), 63.64 ($C_5$′), 38.12 ($C_2$′), 25.91 (C-(*C*H$_3$)$_3$), 18.32 (Si-*C*-(CH$_3$)$_3$), -5.46 (Si-*C*H$_3$) ppm. MS (ESI positive ion mode): m/z calculated for $C_{18}H_{20}N_2O_5$ 376.1271, found 377.1258.

***3'-benzyloxycarbonyl-2'-deoxythymidine-5'-(malic acid dibenzyl ester)-benzyl phosphotriester (20)***

To a solution of **19** (94 mg, 0.25 mmol, 1 eq) in anhydrous dichloromethane (5 mL) kept under argon atmosphere was added a solution of **9** (202 mg, 0.38 mmol, 1.5 eq) in anhydrous dichloromethane (2 mL). A solution of tetrazole in acetonitrile (0.1 M, 2.2 mL, 4 eq.) was added *via* syringe and the resulting clear colourless mixture was stirred for 2.5 hour at room temperature. After completion of the reaction was assessed by TLC, an oxidizing solution of *tert*-butanol peroxide in decane (6.0 M, 0.2 mL, 1 eq.) was added *via* syringe and the mixture was further stirred for 20 min. The mixture was then partitioned between dichloromethane and a saturated solution of sodium chloride. Organic phase were gathered, dried on $Na_2SO_4$, evaporated and co-evaporated with dichloromethane 3 times. 245 mg of yellow oily residue were purified by flash chromatography eluting with a 0-2% gradient methanol in dichloromethane and afforded **20** in small amount (27 mg, 0.032 mmol, 8% yield). Rf($CH_3Cl$/$CH_3OH$:6/1) = 0.90 ≈ Rf($CHCl_3$/$CH_3OH$/$H_2O$:5/3/0.5). $^{1}$H NMR (300 MHz , $CDCl_3$): $\delta$ 8.53 (s, 1H, $H_6$), 7.33 (m, 21H, $H_{Ph}$), 7.32-7.28 (m, 20H, Ph), 6.37 (dd, $^{3}J_{H1'-H2'}$ = 8.88 Hz, 1H, $H_{1'}$), 5.19-4.98 (m, 10H, 4xC$H_2$-Ph, C$H_\alpha$, $H_{3'}$), 4.40-4.13 (m, 3H, $H_{4'}$ and H5'), 3.01-2.93 (dd, 2H, C$H_{2\beta}$), 2.35(m, 1H, $H_{2'}$), 2.03 (m, 1H, $H_{2'}$), 1.72 (s, 3H, $CH_3$) ppm. $^{31}$P NMR (121 MHz, $CDCl_3$): δ -1.86, -1.99 ppm. (Rp and Sp). HRMS (ESI): m/z calculated for $C_{43}H_{43}N_2O_{14}P$: 853.2524, found: 853.2543.

***2'-deoxythymidine-5'-(malic acid)-phosphodiester* (21)**

Phosphotriester **20** (20 mg, 0.023 mmol) was dissolved in methanol (2 mL), palladium on carbon 10% (Degussa® type, 0.002 mmol, 0.3 eq) was added and the resulting suspension was stirred for 24 hours under hydrogen atmosphere (atmospheric pressure) while monitoring by TLC. Upon completion, the suspension was filtered on Celite® 545, and washed several times with methanol. The filtrate was evaporated and 11mg of resulting solid was lyophilized,

affording **3** as a white solid (10 mg, 0.023 mmol, quantitative yield). Rf($CHCl_3$/$CH_3OH$/$H_2O$:5/3/0.5)=0.35. MS (ESI negative mode): m/z calculated for $C_{14}H_{19}N_2O_{12}P$ 438.28, found 438.6. Stability in $D_2O$: 15 minutes.

***Synthesis of 2'-deoxyadenosine-5'-methylhistidyl phosphoramidate***

In a two neck flask, 2'-deoxyadenosine-5'-monophosphoric acid hydrate (200 mg, 0.572 mmoles) and L-histidine methyl ester hydrochloride (979 mg, mmoles, 7 equiv.) were dissolved in a mixture of *tert*-butanol (9 ml) and $H_2O$ (2ml). A few drops of triethylamine ($Et_3N$) were added to facilitate dissolution. An appropriate amount of N,N'-dicyclohexylcarbodiimide (DCC) (826 mg, 4 mmoles, 7 equiv.) was dissolved under argon in 1 ml of *tert*-butanol and was added dropwise to the reaction mixture. The reaction mixture was refluxed carefully for 3-4 hours while stirring under argon. The progress of the reaction was monitored by TLC (*i*PrOH:$NH_3$:$H_2O$). Upon completion, the reaction mixture was cooled down and the solvent was removed by rotary evaporation. The product was isolated by silica column chromatography eluting with $CHCl_3$:MeOH:$H_2O$ gradient and affording the product as a white solid (138 mg, 50% yield). $^1$H NMR (300 MHz , DMSO-$d_6$): δ 8.40 (s, 1H), 8.17 (s, 1H), 8.13 (s, 1H), 7.26 (s, 2H, $NH_2$), 7.05 (s, 1H), 6.35 (apparent t, $^3J_{H1'-H2'}$ = 6.6 Hz, 1H, $H_{1'}$), 4.41 (m, 1H, $H_{3'}$), 4.16 (s, 1H, $H_{4'}$), 3.94 - 3.92 (m, 1H, C*H*COO$CH_3$), 3.83-3.65 (m, 2H, $CH_2$), 3.53 (s, 3H, COOC*H*$_3$), 2.89 (m, 2H, $H_{5'}$), 2.75 -2.61 (m, 1H, $H_{2'}$), 2.33-2.19 (m, 1H, $H_{2'}$). $^{31}$P NMR (121 MHz, DMSO-$d_6$): δ 3.77 ppm. MS (ESI positive ion mode): m/z calculated for $C_{17}H_{23}N_8O_7P$ 482.39, found 483.0.

***2'-deoxyadenosine-5'-histidyl phosphoramidate (2)***

The DCC coupling procedure described for with starting material 2 ′-deoxyadenosine (100 mg, 0.3 mmol), L-histidine dimethyl ester hydrochloride (363 mg, 1.5 mmoles, 5 eq.) and DCC (414 mg, 1.76 mmoles, 7 eq.) afforded **2** as a white solid (85 mg, 0.18 mmol, 61%).

$^1$H NMR (300 MHz , $D_2O$): $\delta$ 8.30 (d, 1H, $^4J_{HIm\text{o-}CH2}$ = 1.5Hz, $H_{Im}$), 8.29 (s, 1H, $H_2$), 8.03 (s, 1$H_8$), 6.95 (s, 1$H_{Im}$), 6.33 (t, $^3J_{H1'-H2'}$ = 6.9 Hz, 1H, $H_{1'}$), 4.57 (bs, 1H, $H_{3'}$), 4.14 (s, 1H, $H_{4'}$), 3.94-3.78 (m, 2H, $H_{5'}$), 3.71-3.59 (m, 1H, C*H*COOH), 2.85-2.64 (m, 3H, $H_{2'}$, C$H_2$), 2.55-2.42 (m, 1H, $H_{2'}$). $^{13}$C (125 MHz, DMSO-d6): $\delta$ 178.29, 155.02, 152.32, 148.42, 139.55, 132.34, 129.21, 118.14, 116.12, 85.67, 83.26, 70.93, 63.59, 55.86, 38.32, 29.34, 23.52 ppm. $^{31}$P NMR (121 MHz, $D_2O$): δ 6.44 ppm. HRMS (ESI): m/z calculated for $C_{16}H_{21}N_8O_7P$ 467.1193, found 467.1182. Stable in $D_2O$ at r.t.

**Phosphate activation strategy:**

***Synthesis of 2'-deoxyadenosine-5'-(β-imidazole lactic acid)-monophosphate (4)***

The imidazolate of 2'-deoxyadenosine-5'-O-monophosphate **16** was synthesized as described in Chapter II. A solution of **16** (20 mg, 50 µmoles), β-imidazole lactic acid (15.6 mg, 100 µmoles, 2 equiv.), zinc chloride (7 mg, 50 µmoles, 1 equiv.) in *N*-ethylmorpholine aqueous buffer 0.2 M (2 mL, pH 7.5) was let standing at room temperature for 5-20 hours. The reaction was monitored by $^{31}$P NMR spectroscopy or ion exchange HPLC, quenched with a 0.25M EDTA (ethylenediaminetetraacetic acid) solution in order to break-down the nucleotide-metal complex. The product was isolated after HPLC purification on an ion exchange column (Source 15Q, Amersham Biosciences) by running a gradient of triethylammonium bicarbonate (TEAB) 1M in Millipore water. Corresponding fractions were analyzed by ESI-MS prior to lyophilisation. Removal of the triethylammonium salt was completed by two additional lyophilisation steps, affording **4** as a white solid. $^{1}$H NMR (300 MHz, $D_2O$, 25°C): $\delta$ 8.4 (s, 1H, $H_8$), 8.3 (s, 1H, $H_{Im}$), 8.1 (s, 1H, $H_2$), 7.0 (s, 1H, $H_{Im}$), 6.4 (apparent t, $^{3}J_{H1'-H2'}$ = 6.8 Hz, 1H, $H_{1'}$), 4.5 (m, 1H, $H_{3'}$), 4.2 (m, 1H, $H_{4'}$), 4.1 (s, 2H, $CH_{2\beta}$), 3.9 (m, 2H, $H_{5'}$), 2.8 (m, 1H, $CH_{\alpha}$), 2.7 (m, 1H, $H_{2'}$), 2.5 (m, 1H, $H_{2'}$) ppm. $^{13}$C NMR (125 MHz, $D_2O$): $\delta$168.5, 155.8, 152.8, 148.9, 140.2, 133.9, 125.8, 118.8, 86.5, 84.0, 71.8, 64.6, 54.3, 52.6, 39.5 ppm. $^{31}$P NMR (121 MHz, $D_2O$): $\delta$ -0.89 ppm. HRMS (ESI, negative ion mode): m/z calculated for $C_{16}H_{19}N_7O_8P^-$: 468.1038, found 468.1030. Stability in $D_2O$: 20% hydrolysis after 12h. Stability in RTRB: 30 minutes.

## III.4.3 Enzymatic reactions

**DNA polymerase reactions:** All reactions were duplicated; tripliquets were realized in case of conflicting results and in order to verify the reproducibility of the results. End-labeled primer was annealed to the suited template by combining primer and template in a 1:2 molar ratio and heating the mixture at 70°C for 10 min followed by slow cooling to room temperature over a period of 1.5 h. Since a risk of early departure of the leaving group was experimentally observed in the presence of a divalent cation with malic acid-dAMP, the sample preparation protocol for single nucleotide incorporation assay and for steady-state kinetic studies was modified in order to preserve the triphosphate mimic from being

prematurely degraded by the magnesium ions present in the Reverse Transcriptase Reaction Buffer (RTRB).

A solution of substrate was pre-incubated at 37°C for 1 minute and mixed with a 37°C-preincubated mixture of enzyme, appropriate primer template hybrid and RTRB (10 µL reaction volume). For the incorporation a series of 10 µL-batch reactions was performed with the enzyme HIV-1 RT (Ambion, 10 U/µL stock solution, specific activity 11.700 U/mg, concentration 0.85 mg/mL). The final mixture contained 125 nM primer template complex, RTRB (250 mM Tris.HCl, 250 mM KCl, 50 mM $MgCl_2$, 2.5 mM spermidine, 50 mM dithiothreitol (DTT); pH 8.3), 0.025 U/µL (32.1 nM) HIV-1 RT , and different concentrations of phosphoramidate or phosphodiester building blocks (0.1-1 mM). In the control reaction with the natural nucleotide, a 10 µM dATP concentration was used. The mixture was incubated at 37°C and aliquots were quenched after 5, 10, 20, 30 and 60 min. For elongation experiments, the same mixture with appropriate primer:template hybrid was incubated at 37°C and aliquots were drawn after 30, 60 90 and 120 min. In the control reaction, 50 µM dATP was used. Aliquots (1.5 µL) were drawn at 5 time points and quenched for 8 min at 85°C in formamide buffer (6µL).

**Steady-state kinetics of single nucleotide incorporation:** The steady-state kinetics of single nucleotide incorporation of L-His-dAMP **1** and ILA-dAMP **2** and of the natural nucleoside triphosphate (dATP) was determined by gel-based polymerase assay. In all the experiments, the template T1 and the primer P1 were used. The primer and template in 1:2 molar ratio were hybridised in a buffer containing 20 mM Tris.HCl, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.3 and used in an amount to provide 125 nM concentration of the primer in each 10 µL reaction. The range of concentrations for substrate was optimized according to a $K_M$ value for the incorporation of an individual nucleotide (0.02 mM - 1 mM for **1** and **2**, 0.01 to 0.1 mM for dATP). Reaction mixtures containing the HIV-1 RT enzyme (Ambion, 10 U/µL) in appropriate concentration to attain 5-20 % conversion (0.0063 U/µL for **1** and dATP, 0.025 U/µL for **2**) were incubated at 37°C and aliquot were drawn at 1, 2, 3, 30 and 60 minutes. The reactions were quenched by addition of the buffer 80% formamide, 2 mM EDTA, 1X TBE buffer. The analysis of polymerase reaction was performed by polyacrylamide gel electrophoresis. The initial incorporation rates ($V_0$) were calculated based on the percentage of the extension production (P+1 band). The kinetic parameters ($V_{Max}$ and $K_M$ were determined by plotting $V_0$ (nM/min$^{-1}$) versus substrate concentration (mM) and fitting the data to a non-linear Michaelis-Menten regression using GraphPad Prism Software (Version 5.0).

# IV. Chapter four: Iminodiacetic-phosphoramidates as metabolic prototypes for diversifying nucleic acid polymerization *in vivo*

The results reported in this chapter were published in *Nucleic Acid Research* in 2010 under the title 'Iminodiacetic-phosphoramidates as metabolic prototypes for diversifying nucleic acid polymerization *in vivo'*. Giraut, A., Xiaoping, S., Froeyen, M., Marlière, P., Herdewijn, P., *Nucleic Acid Res.,* **38**, 2541-2550.

## *IV.1 Introduction*

The aim of our project is to develop and introduce additional types of nucleic acids in living cells. As a long term objective, these nucleic acids could be used to avert genetic pollution by preventing dissemination of artificial hereditary messages in natural ecosystems, including the human body (56). In formal terms, this project amounts to the systematic search of the chemical space in order to identify an assortment of activated monomers carrying a set of complementary bases. Another aspect of this project is to discover nucleic acid enzymes that would use activated monomers as substrates in polymerization, ligation or recombination. The result of this interplay would be templated-reproduction of polymers which would convey additional genetic information but not interfere with RNA and DNA metabolism or function.

Base and backbone variants of DNA and RNA are the topic of intense research in several laboratories, auguring well for the feasibility of xenonucleic acid (XNA) reproduction *in vivo* (60;68). It is our strong belief that unnatural leaving groups used for activation of XNA are crucial propagating additional types of nucleic acids in living cells. Indeed, the chemical structure of the leaving group can be elaborated leading to segregation polymerization of XNA polymers from that of indigenous DNA and RNA monomers and establishing an informational enclave. Another result, as a consequence, is disentangling the XNA polymerization from the phosphoanhydride economy of the cell and thus establishing an energetic enclave as well. The dual role of nucleoside triphosphates as both energetic currency and precursors for the informational polymers RNA and DNA, drastically limits the range of metabolic and genetic reprogramming that can be accomplished in living cells. The

judicious implementation of novel leaving groups in metabolism could overcome the need for physical compartmentalization of unnatural information transfers and enable the launching and sustaining of autonomous hereditary procedures *in vivo*.

In the aim of broadening our search of the chemical space, we addressed the synthesis and enzymatic assay of the deoxyadenylate adduct of iminodiacetate. Although this isomer of aspartate is not commonly found in living cells, its biosynthesis and recycling from glycine and glyoxylate could be readily implemented in metabolism. Formation of the prototype metabolite triphospho-imino-diacetate (TPI) and its reaction with nucleosides can also be experimentally studied to investigate novel energetic and informational metabolism and its potential implementation in bacterial cells.

The enzymatic synthesis of nucleic acids makes use of nucleoside triphosphates as substrates and this process is driven by the release and hydrolysis of pyrophosphate. The selection of new leaving groups for the enzymatic synthesis of nucleic acids should be accompanied by a selection of appropriate polymerases which can use the modified nucleotides as building blocks for gene synthesis, independent of the cellular gene-synthesis machinery (56). Since HIV-1 RT shows broad substrate specificity, this enzyme was chosen as a primary polymerase for selection of potential leaving groups to activate nucleoside triphosphate mimics.

Despite the fact that many polymerases as HIV-1 reverse transcriptase, Taq DNA polymerase, Vent (exo-) DNA polymerase , are able to process selected substrates like amino acid dAMP phosphoramidates (82), HIV-1 RT was chosen as the preferred enzyme. In our study the pyrophosphate leaving group (linked to a deoxynucleotide in Figure IV.1a) is replaced by a completely different molecule with potential chelating properties. It has been indicated that the L-aspartic acid derivatives of dATP (Figure IV.1b), dGTP, dCTP and dTTP demonstrated conserved single nucleotide incorporation and limited elongation properties (83). However incorporation of L-His-dAMP by HIV-1 RT proceeded with lower efficiency. In contrast to its congener L-Asp-dAMP, primer chain elongation was slightly improved when L-His-dAMP was used (82). It is important to note that regardless of which amino acid was used, DNA synthesis dramatically slowed down after incorporation of two nucleotides. Understanding and resolving this stalling issue is a prerequisite for the future use of such modified nucleotides for enzyme-catalyzed DNA synthesis *in vivo*.

A model postulated by Steitz explains the way by which the incoming deoxynucleotide triphosphate (dNTP) is bound in the active site of the polymerase for incorporation in the

growing DNA chain (12). This model takes into account the importance of the chelation of two divalent magnesium ions required in the polymerization mechanism which are thought to bring the free 3'-OH group of the primer for in-line attack the α-phosphate group of dNTP. More recently, Garforth *et al.* suggested that enzyme processing could occur when only two potential chelating moieties are present, since in their studies a deoxynucleoside diphosphate served as substrate for HIV-RT and resulted in DNA strand elongation (84).

In the case of phosphoramidate substrates, the necessity of the leaving group to carry two potential chelating moieties has been acknowledged in chapter II. To broaden the chemical space and get further insight in the structure-activity relationship, we altered the substitution pattern of the nitrogen atom of the phosphoramidate moiety, while a) taking care of potential chelating properties and b) focusing on potential metabolic accessibility of the new leaving group. Enhanced coordination of the magnesium ion by the leaving group moiety of the modified nucleotide might influence incorporation kinetics and efficiency of the chain elongation reaction. In addition, iminodiacetic acid (Figure IV.1c) can be considered as potentially metabolically accessible and potentially prone to catabolism into non-toxic cellular constituents (i.e. glycine and a two-carbon atom fragment glyoxalic acid). As a consequence, this degradation reaction may have a driving effect on the equilibrium of the nucleotide incorporation reaction.

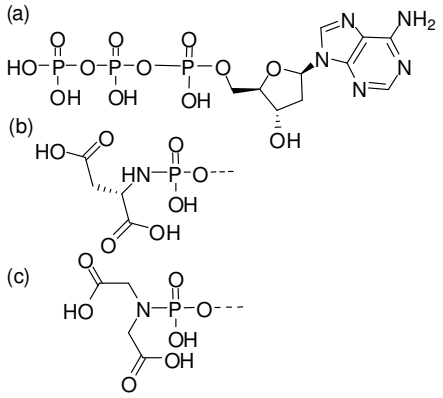



**Figure IV.1:** Molecular structures of deoxyadenosine (a) -triphosphate, (b) -aspartic acid phosphoramidate, (c) –iminodiacetic acid phosphoramidate.

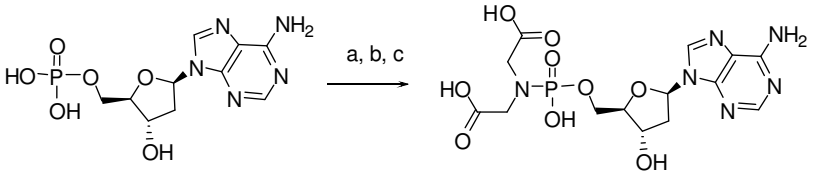
The iminodiacetate dAMP phosphoramidate (IDA-dAMP, Figure 1c) possesses no chiral centre, ensuring less stereochemical constraint than for the chiral L-aspartic acid. Furthermore, the diacetic acid substituted amine is present in the molecular structure of ethylenede diamine tetraacetic acid (EDTA), known for its great chelating properties towards

a diversity of metal ions. Finally, it has been shown recently that protonation of the pyrophosphate leaving group is important and facilitated by a neighbouring general acidic residue in the active site (often a lysine residue) in all four classes of nucleic acid polymerases (89). A departing amide ion is a very strong base as illustrated by the high pKa of sodium amide (pKa in ammonia solution >30) (109) and should therefore undergo extremely rapid protonation.

## *IV.2. Results and discussion*

### IV.2.1 Synthesis of Iminodiacetate dAMP phosphoramidate (IDA-dAMP)

The synthesis of the methyl ester of IDA-dAMP was accomplished according to the method described by Wagner and colleagues starting from the corresponding nucleoside monophosphate (75). Deprotection of the methyl ester was carried out with sodium hydroxide in methanol-water solution. Scheme IV.1 shows the synthetic route, which is an easy-to-perform two-step reaction.




**Scheme IV.1:** Preparation of iminodiacetic acid dAMP phosphoramidate: a) iminodiacetic acid, DCC, *tert*-butanol, water; b) sodium hydroxide, methanol, water (1M); c) TEAB 1M.

### IV.2.2 Single Nucleotide Incorporation

HIV-1 reverse transcriptase serves as a catalyst in the HIV-1 viral replication process and uses deoxynucleotides as substrates. This polymerase is error-prone and thus possesses a high mutation rate (4;18). Previous experiments carried out with the L-aspartic acid phosphoramidate of dAMP demonstrated that this amino acid was an acceptable leaving group for the polymerase in the nucleotidyl incorporation process (81;82). In the present study, we evaluated the capacity of HIV-1 RT to incorporate IDA-dAMP as a substrate in the primer-template complex P1:T1. The initial experiments were carried out using a template with an overhang of one thymidine nucleotide followed by three non-thymidine bases (Table

IV.1). Incorporation efficiency was analyzed by the polyacrylamide gel-based single nucleotide incorporation assay (110).

**Table IV.1**. Overview of the primer-template complexes used in the DNA polymerase reactions. Bold letters indicate the template overhang in the hybridised primer-template duplex.

| | |
|---|---|
| Single nucleotide incorporation and kinetic experiments | |
| P1 | 5'-CAGGAAACAGCTATGAC-3' |
| T1 | 3'-GTCCTTTGTCGATACTG**TCCC**-5' |
| Elongation experiments | |
| P1 | 5'-CAGGAAACAGCTATGAC-3' |
| T2 | 3'-GTCCTTTGTCGATACTG**TTTTTTT**-5' |
| T3 | 3'-GTCCTTTGTCGATACTG**TTTTTTTGGAC**-5' |

The phosphoramidate analogue IDA-dAMP was correctly processed by HIV-1 RT converting a primer to a (P+1) strand with a yield up to 86% in 60 min, at 500 µM deoxynucleotide concentration (Figure IV.2). So as for L-Asp-dAMP (50% P+1 product at 100 µM), efficient substrate incorporation was also detected with a 5-fold lower substrate concentration (40% P+1 product at 100 µM).

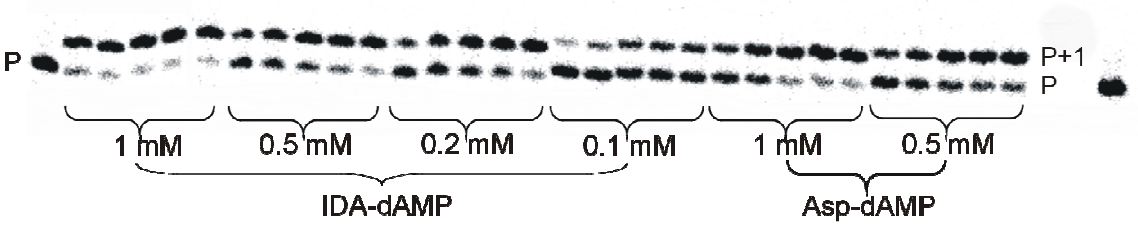



**Figure IV.2:** Profile of incorporation of dAMP into P1:T1 by HIV-1 reverse transcriptase using IDA-dAMP as substrate. Aliquots were drawn at 10, 20 30, 60 and 120 minutes.

### IV.2.3 Elongation experiments

Previous studies carried out with L-Asp-dAMP and HIV-1 RT showed its ability to serve as a substrate for template dependent incorporation of more than one phosphoramidate nucleotide. Primer elongation of up to 6 nucleotides was observed only to a limited extent, and with prevalence of the (P+2) and (P+3) products. Such a stalling represents a significant hindrance for a future use of phosphoramidate nucleotides for enzymatic DNA synthesis *in vivo*. For the

evaluation of the strand elongation capacity of IDA-dAMP using the same enzyme, the primer-template duplexes P1:T2 and P1:T3 were used (Table IV.1).

A range of concentrations of IDA-dAMP was incubated at the appropriate temperature with the primer-template complex and 0.025 U/µL of enzyme. Reaction aliquots were drawn between 30 and 120 minutes time points and analyzed by 20% polyacrylamide gel electrophoresis. In the P1:T2 experiment, the template possesses an overhang of seven thymidine nucleotides. Using IDA-dAMP as a substrate, we observed a prevalence of the (P+6) product, indicating an excellent elongation capacity (Figure IV.3). The initial stalling at the (P+2) and (P+3) products disappears over time and is not so pronounced as with L-Asp as a leaving group (Numerical data are gathered in Table IV.2)**.** However, we did not observe formation of a P+7 when P1:T2 system was used.

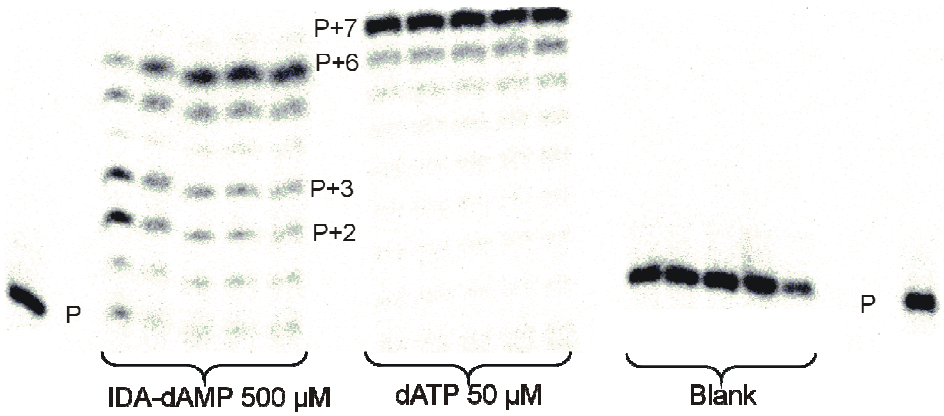



**Figure IV.3:** Profile of elongation of P1:T2 with dAMP by HIV-1 RT using IDA-dAMP as substrate. Aliquots were drawn at 10, 20 30, 60 and 120 minutes.

The major product of L-Asp-dAMP incorporation is the (P+2) resulting oligonucleotide, whereas the (P+6) product is only present in trace amounts and barely quantifiable. In contrast, IDA-dAMP incorporation results mostly in formation of the (P+6) oligonucleotide (Table IV.2).

**Table IV.2. L-Asp-dAMP *versus* IDA-dAMP in the elongation of P1 directed by template T2: % of P+n product after 120 min reaction**

| product | % substrate[a] L-Asp-dAMP | % substrate[a] IDA-dAMP |
|---|---|---|
| P+6 | Traces | 40 |
| P+5 | Traces | 19 |
| P+4 | Traces | 3 |
| P+3 | 34 | 14 |
| P+2 | 64 | 12 |

[a] % of the total amount of radioemitting oligonucleotides in the mixture

In the subsequent P1:T3 experiment using a P1:T3 duplex, where the seven thymidine overhang of the template was flanked by four non-thymidine units, the primer extension resulted in formation of a (P+7) product (Figure IV.4).

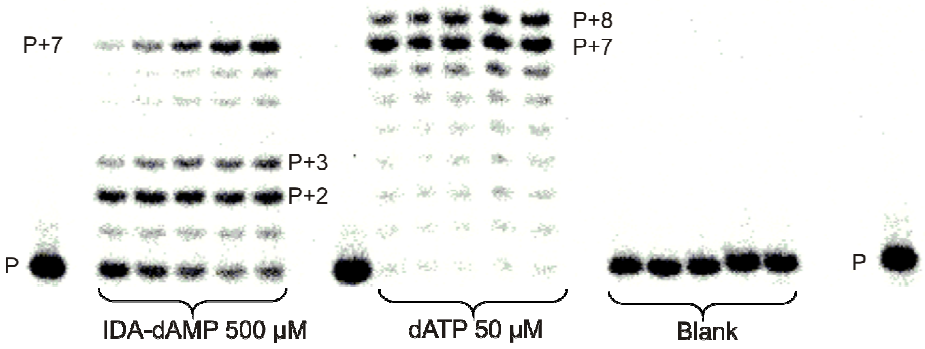



**Figure IV.4:** Profile of elongation of P1:T3 with dAMP by HIV-1 RT using IDA-dAMP as substrate. Aliquots were drawn at 10, 20 30, 60 and 120 minutes.

The observed full-length elongation obtained when using the purine-elongated template T3, could be due to the increased affinity of the longer template overhang for the enzyme and the formation of a more stable tertiary complex. It has been shown by Kohlstaedt in 1992 that the template strand binds to the fingers subdomain of the polymerase active site (91). This was

confirmed in the crystal structures observed by Huang for HIV-1 RT and by Doublié for T7 DNA polymerase (14;17). In addition, the binding affinity of RT with DNA duplexes as template is increased when template overhang was extended to 6 additional bases (111). According to models based on crystal structures of the complex (111;112), the β3-β4 loop situated in this finger domain contacts the template three nucleotides upstream from the primer terminus (near or at $Leu^{74}$). Kew *et al.* postulated that a stronger complex between the fingers subdomain of a HIV-RT and the template overhang may be responsible for higher processivity by the enzyme (113). Although it seems that binding and incorporation of a natural triphosphate substrate dATP is not dependent on the length of the template overhang (as observed in Figure IV.3 and 4 and in (112), it appears to be the case with the triphosphate mimic IDA-dAMP. The presence of a four base 5'-overhang could favour a more stable dsDNA-protein complex. The template length also seems to influence the ability of HIV-1 RT to recognise and incorporate an incoming dNTP carrying a modified leaving group.

The superior substrate properties of IDA-dAMP as compared to L-Asp-dAMP might also be arising from the better chelating ability of the *N*-diacetate group of IDA-dAMP, from structural differences of the transition state, or from modified conformational changes of the enzyme. In the case of T7 RNA polymerase, Yin and Steitz proposed a model where cross-linking between the DNA helix and the enzyme depends on the presence of pyrophosphate. They demonstrated that the nature of the electrostatic interactions of this complex undergo changes between the pre- and post-translocation state of nucleotide incorporation (108). In a previous model postulated by Steitz, one magnesium ion is indeed coordinated by the 3′-end of the primer and the α-phosphate, while the second magnesium ion is thought to facilitate the leaving group properties of the pyrophosphate moiety by forming interactions with the β- and γ-phosphates of the nucleotide (12). It can be hypothetized that the better stability of the tertiary complex compensates for the loss of stability induces by the replacement of PPi.

### IV.2.4 Kinetic experiments

Further investigation of IDA-dAMP was focused on kinetics of incorporation. Kinetic parameters for the incorporation of both the natural and the modified substrate by HIV-1 RT were determined on the basis of the single completed hit model (88;114). In the study, P1 and T1 (Table IV.1) were used as the primer and template DNA strands, respectively. The steady-state kinetic values $K_M$ and $V_{Max}$ corresponding to the substrate efficiency of the nucleotide

triphosphate analogues are reported in Table IV.3. Along with showing a similar single incorporation after 60 minutes and far better elongation results, IDA-dAMP also demonstrates better incorporation kinetics than L-Asp-dAMP for HIV-1 RT. The measured $V_{Max}/K_M$ ratio towards HIV-1 RT is only one order of magnitude lower than that of dATP. In comparison, this ratio was for L-Asp-dAMP $10^3$ fold lower than the one of the natural substrate. Similarly to the L-Asp phosphoramidate analogue, the $K_M$ of IDA-dAMP is much higher than for the natural substrate, whereas the $V_{Max}$ is 1.3 fold lower. The increased capacity for chain elongation is paralleled by a similar increase in incorporation efficiency. It is clear that enzyme evolution should focus on the selection of a polymerase enhancing $K_M$ rather than $V_{Max}$, which is easier to achieve than the inverse.

**Table IV.3.** Determination of the kinetic parameters of the incorporation of dAMP with the natural pyrophosphate leaving group (dATP) and the iminodiacetate leaving group (IDA-dAMP) into $P_1T_1$ by HIV-1 RT DNA polymerase at 0.025 U/µL.

| | $K_M$ [µM] | $V_{max}$ [nM.min$^{-1}$] | $V_{max}/K_M$ [min$^{-1}$] |
|---|---|---|---|
| dATP | 29 ± 5 | 65.7 ± 4.3 | $2.26x10^{-3}$ |
| IDA-dAMP | 312.5 ± 70.5 | 49.0 ± 4.5 | $0.16x10^{-3}$ |

## IV.2.5 Model of IDA-dAMP bound to HIV-1 reverse transcriptase

With the IDA-dAMP molecule bound to reverse transcriptase, stable molecular dynamics (MD) trajectories were obtained. In the calculated ground state complex (Figure 5b), the spatial arrangement of two $Mg^{2+}$ ions is comparable to their situation in the original TTP complex (Figure 5a) (17). Although in the ground state the carboxylate groups of IDA-dAMP (mimicking the β and γ phosphate groups) are not yet involved in an ionic bond with the $Mg^{2+}$ ions, they are stabilised by ionic interactions with Lys65A, Arg72A and Gln151A. The difference in the binding affinity between PPi and IDA could also contribute to a difference in kinetics or to the lower processivity of the enzyme when challenged with IDA-dAMP in place of dATP (92). In the ground state crystal structure obtained with dTTP, Gln151 only contacts the leaving group of the dTTP, without forming an actual ionic bond.

The divalent magnesium ions are located in a very strong network of ionic bonds which involves the non-bridging oxygen of the phosphoramidate group and the three aspartate residues Asp185A, Asp110A and Asp186A, which are highly conserved in the polymerase

family (111;115). The ionic bond interactions observed between the divalent cation and residue Asp186A in the ground state is specifically noted in the complex of HIV-1 RT with both IDA-dAMP and Asp-dAMP (82;93).

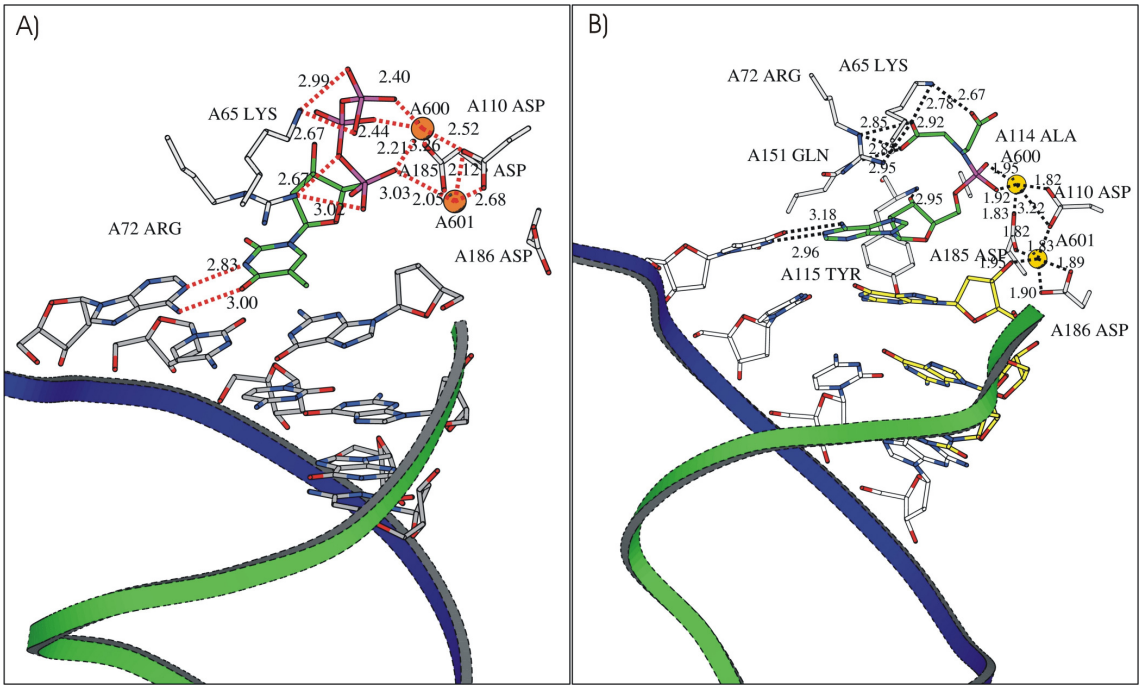
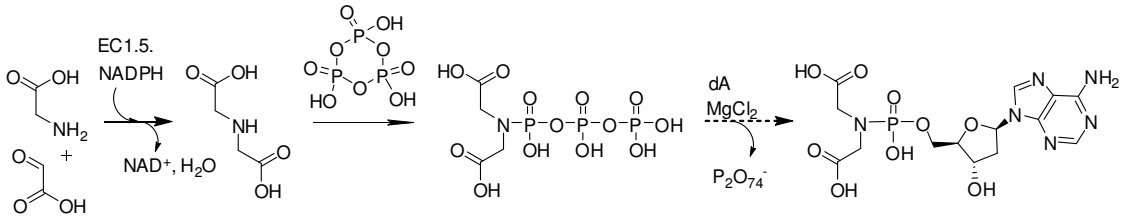



**Figure IV.5:** A) The thymidine triphosphate (TTP) in the RT dNTP pocket (X-ray structure from Huang 1998 (17)). The primer strand is shown as a green ribbon; the template strand is in blue. The stabilizing interactions of the TTP within the complex are shown. Hydrogen bond distances between the bases are indicated in Å. The ligation of the 2 $Mg^{2+}$ ions to the complex is also visualized by dashes and distances in Å. B) Average structure of IDA-dAMP in the RT dNTP binding pocket. This figure was generated using Bobscript, Molscript and Raster3d (116-118).

## IV.2.6 Metabolic accessibility

In order to investigate the metabolic accessibility and recycling potential of a novel triphosphate mimic, we synthesized triphospho-iminodiacetate (TPI) and tested the possibility of displacement of the pyrophosphate moiety by deoxyadenosine (Scheme IV.2). The triphosphate was formed by reacting trimetaphosphate with iminodiacetic acid, adapting a method described by Feldmann *et al.* (119). However, the pyrophosphate displacement reaction with deoxyadenosine in the presence of magnesium ions with release of

pyrophosphate was not successful due to poor nucleophilicity of the 5′-hydroxyl group of deoxyadenosine and solubility problems. On the other hand, cleavage of cyclic trimetaphosphate by adenosine in presence of $MgCl_2$ in neutral conditions was previously observed by Yamagata (120). The presence of a divalent cation which would facilitate the pyrophosphate displacement implies the use of a non-nucleophilic polar solvent. Therefore, further experiments are needed to circumvent solubility problems. Another approach to solve the aforementioned difficulty with pyrophosphate displacement consists in developing biochemical agents capable of catalyzing the reaction.




**Scheme IV.2:** Potential metabolic availability of IDA-dAMP using 1) enzymatic condensation of simple carbon fragments by a dehydrogenase (EC.1.5), 2) cyclotrimetaphosphate as triphosphate precursor and 3) pyrophosphate displacement by deoxyadenosine.

## *IV.3 Conclusion*

Deoxyadenosine triphosphate analogues in which the pyrophosphate moiety is replaced by L-aspartic acid had previously been shown to act as substrates for DNA synthesis catalyzed by HIV-1 RT. The use of iminodiacetate dAMP as a potential substrate improved both incorporation and elongation results. The contribution of the pyrophosphate leaving group to DNA synthesis, in terms of catalytic efficiency and fidelity of DNA polymerase β, was studied recently (70). These data suggest a leaving-group-induced change at the rate limiting step due to stabilisation of the pyrophosphate group release. It is clear that the substitution of pyrophosphate by L-Asp and IDA moieties would be likely to have a major impact on this process. Furthermore, chemical modification of the leaving group influences active-site structural differences between correct and incorrect base-paired transition state (71).

Elongation results with IDA-dAMP allowed us to confirm the positive influence of a longer template overhang for the recognition and incorporation of several successive building blocks carrying a modified leaving group. Therefore, our results validated the importance of

interactions between finger subdomain residues and template's upstream nucleotides during the HIV-1 RT directed DNA synthesis.

Modelling experiments showed us that in the ground state the carboxylate groups are most likely to be bound to an intricate network of amino acid residues, Lys65, Arg72 and Gln151, located in the enzyme's active site *via* divalent cations. Our study suggests that in the ground state the magnesium ion could also form an ionic bond with Asp110, Asp185 and Asp186. The last interaction mentioned (Asp186) is observed only in the case of IDA-dAMP and L-Asp-dAMP. Therefore this model implies that in the ground state the stabilisation of IDA-dAMP is enhanced and involves 6 amino acid residues instead of 4 with the natural substrate and 5 with Asp-dAMP (82). This increase in amino acids interactions observed in the complex formation with IDA-dAMP is reflected in a change in $K_M$. The electrostatic involvement of residues Gln155A and Asp186A has not been observed in previous studies. The difference in binding affinity between pyrophosphate and iminodiacetic acid is likely to contribute to a difference in kinetics (92). In addition, the difference in intrinsic energy of the released leaving group could be a possible "brake" for further conformational changes of the enzyme (92;108).

The nucleoside phosphoramidates developed in this study lend themselves to metabolic implementation, as described in Scheme IV.2. With the help of the divalent magnesium cation, a further nucleophilic attack of the α-phosphate by the 5′-hydroxyl group of adenosine should form the phosphoramidate IDA-dAMP with the energetically favoured release of a pyrophosphate. The biosynthesis would start from the two common metabolites glycine and glyoxylate, which is its corresponding 2-oxo-acid. A specific transaminase, namely serine-glyoxylate aminotransferase) performs interconversion of glycine and glyoxylate in certain methylotrophic bacteria (121). Reductive condensation of glycine with glyoxylate resulting in iminodiacetate is reminiscent of octopin biosynthesis. Enzymes directing the production of these nutrient shuttles could therefore be harnessed for our purpose (122;123)**.** Furthermore, ammonia reacts in neutral aqueous media with cyclotriphosphate to yield triphosphoramidates and the equilibrium constant of this reaction is about 1.52 $min^{-1}$ (119). We propose that this reaction could be readily observed in the case of the secondary amine IDA, spontaneously providing TPI in good yield (80 % yield after 24 hours at room temperature). Putative enzymes capable of accelerating this spontaneous reaction could possibly be found in the class of hydrolases that act on acid anhydrides, such as trimetaphosphatase (EC 3.6.1.2) (124).

Nevertheless, our study showed that further condensation of TPI with the 5'-hydroxyl of the nucleoside deoxyadenosine did not proceed spontaneously under the various conditions tested. Since this reaction features the nucleophilic attack of a phosphoanhydride by a primary alcohol as performed by many kinases, the systematic assay of such enzymes can unveil a biocatalyst marginally endowed with the desired activity.

## *IV.4. Experimental section*

### IV.4.1 General methods

NMR spectral analyses were carried out and processed as described in Chapter II. Chemical shifts are expressed in parts per million (*ppm*) by frequency. $^{1}H$ chemical shifts are referenced to an internal TMS signal ($\delta$ = 0.00 ppm), $^{31}P$ NMR chemical shifts are referenced to an external 85 % $H_3PO_4$ standard ($\delta$ = 0.00 ppm). Standard and exact mass spectra were measured as described in Chapter II. Chemicals of analytical or synthetic grade were obtained from commercial sources and were used as received (deoxyadenosine monophosphate: Sigma Aldrich; DCC, dimethyl iminodiacetic acid hydrochloride: Fluka; *tert*-butanol and triethylamine: Acros). Technical solvents were obtained from Brenntag (Deerlijk, Belgium). Acetonitrile HPLC Grade was purchased from Fischer Scientific. Flash silica chromatography was performed on Davisil® silica gel 60, 0.040–0.063 mm (Grace Davison). Thin Layer Chromatography was performed on Alugram® silica gel UV254 mesh 60, 0.20 mm (Macherey-Nagel).

**DNA duplexes preparation and polyacrylamide gel-based electrophoresis** were carried out as described in Chapter II.

### IV.4.2 Synthesis

***Synthesis of 2'-deoxyadenosine-5'-(dimethyl iminodiacetate) phosphoramidate (Intermediate for 10)***

In a two-neck flask, 2′-deoxyadenosine-5′-monophosphoric acid hydrate (100mg, 0.286 mmoles) and dimethyl iminodiacetate hydrochloride (283 mg, 1.430 mmoles, 5 equiv.) were suspended in a mixture of 1, 4-dioxane (9ml) and *N,N*-dimethylformamide (1ml). A few drops of triethylamine were added to the solution to facilitate dissolution. Then, a solution of DCC (414 mg, 2.004 mmoles, 7 equiv.) in 1,4-dioxane (1 ml) was added and the reaction mixture

was heated for 3 hours while stirring under nitrogen atmosphere. The progress of the reaction was monitored by TLC ($CHCl_3$:$CH_3OH$:$H_2O$ 5:3:0.5). Upon completion, the reaction mixture was cooled and the solvent was removed by rotary evaporation. The residue was resuspended in water (15 ml), extracted with diethyl ether (3×10 ml), and the aqueous phase was lyophilised. The resulting solid was subjected to column chromatography on silica gel using the following solvent gradient: $CHCl_3$:$CH_3OH$:$H_2O$ (5:1:0; 5:2:0.25; 5:3:0.25; 5:3:0.5). The product was obtained as a white solid (58mg, 43%).

$^1$H NMR (500 MHz, $D_2O$): $\delta$8.46 (s, 1H, $H_8$), 8.24 (s, 1H, $H_2$), 6.50 (apparent t, 1H, $^3J_{H1'-H2'}$ = 6.5Hz, $H_{1'}$), 4.72 (m, 1H, $H_{3'}$), 4.22 (m, 1H, $H_{4'}$), 4.00 (m, 2H, $H_{5'}$), 3.78 (s, 1H, $CH_2COOCH_3$), 3.69 (s, 1H, $CH_2COOCH_3$), 3.56 (s, 6H, $CH_3$), 2.89 (m, 1H, $H_{2'a}$), 2.61 (m, 1H, $H_{2'b}$) ppm. $^{13}$C NMR (125 MHz, $D_2O$): $\delta$174.27, 156.07, 153.21, 149.36, 140.38, 119.14, 86.63, 84.12, 72.13, 64.67, 52.62, 49.28, 39.25 ppm. $^{31}$P NMR (121 MHz, $D_2O$): $\delta$6.67 ppm. HRMS (ESI): calculated for $C_{16}H_{22}N_6O_9P$ 473.1186, found: 473.1180.

***Synthesis of 2'-deoxyadenosine-5'-iminodiacetate phosphoramidate (10)***

A solution of 2'-deoxyadenosine-5'-dimethyl iminodiacetate phosphoramidate (50mg, 0.105 mmoles) in 0.4 M sodium hydroxide solution in MeOH:$H_2O$ (4:1) (2mL), was stirred at room temperature under nitrogen for 2 hours. The progress of the reaction was monitored by TLC (*i*PrOH:$H_2O$:$NH_3$ 7:1:2). Upon completion, the reaction mixture was neutralised by addition of 1M triethylammonium bicarbonate. The solvent was removed under reduced pressure. The residue was purified by column chromatography and eluted with the following solvent gradient: $CHCl_3$:$CH_3OH$:$H_2O$ (5:1:0 to 5:4:1). The product was isolated and concentrated, affording a white solid (31mg, 65% yield).

$^1$H NMR (500 MHz, $D_2O$): $\delta$8.46 (s, 1H, $H_8$), 8.25 (s, 1H, $H_2$), 6.50 (apparent t, 1H, $^3J_{H1'-H2'}$ = 6.48Hz, $H_{1'}$), 4.70 (m, 1H, $H_{3'}$), 4.26 (m, 1H, $H_{4'}$), 4.02 (m, 2H, $H_{5'}$), 3.64 (m, 4H, $CH_2COOH$), 2.83 (m, 1H, $H_{2'a}$), 2.59 (m, 1H, $H_{2'b}$) ppm. $^{13}$C NMR (125 MHz, $D_2O$): $\delta$ 171.96, 156.26, 153.35, 149.49, 140.68, 119.41, 86.62, 84.42, 71.87, 64.67, 49.67, 39.62 ppm. $^{31}$P NMR (121 MHz, $D_2O$): $\delta$ 8.34 ppm. HRMS (ESI negative ion mode): calculated for $C_{14}H_{19}N_6O_9P$ 446.0951, found: 445.0864. Stability in $D_2O$ at r.t.: superior as 72h.

## IV.4.3 Metabolic availability

Trisodium trimetaphosphate (153 mg, 0.5 mmoles) and iminodiacetic acid disodium salt (979 mg, 5 mmoles, 10 equiv.) were dissolved in double distilled water (6 mL) and the resulting clear mixture was let at r.t. for 1 day and monitored by $^{31}$P NMR (121 Hz, 90% $H_2O$+10% $D_2O$). Simultaneously with disappearance of the starting material (-20.79 ppm, s), formation of three news peaks (-0.39 ppm (d, $J_{P\text{-}P}$ = 20.06 Hz), -4.97 ppm (d, $J_{P\text{-}P}$ = 17.58 Hz), -19.65 ppm (t, $J_{P\text{-}P}$ = 18.99 Hz)), assigned to iminodiacetic acid triphosphate (TPI), was observed (80% yield, from $^{31}$P peak integration). The mixture was lyophilised; solvent (DMF, DMSO, hexamethylphosphoramide, or mixture thereof with water) (1mL), deoxyadenosine (16 mg, 0.0625 mmoles) and magnesium chloride (127 mg, 0. 625 mmoles) were added. The resulting mixture was stirred at r.t. for 1 hour and monitored by $^{31}$P NMR. Disappearance of TPI signals was simultaneous with appearance of two singlets ($\delta$ 0.38 ppm and -9.24 ppm), attributed respectively to iminodiacetic monophosphate and to inorganic pyrophosphate.

## IV.4.4 Enzymatic reactions

**DNA polymerase reactions using IDA-dAMP as substrate**

All reactions were duplicated; tripliquets were realized in case of conflicting results and in order to verify the reproducibility of the results. For the incorporation of IDA-dAMP a series of 20 µL-batch reactions was performed with the enzyme HIV-1 RT (Ambion, 10 U/µL stock solution, specific activity 8.095 U/mg). The final mixture contained appropriate 125 nM primer template complex, RT buffer (250 mM Tris.HCl, 250 mM KCl, 50 mM $MgCl_2$, 2.5 mM spermidine, 50 mM dithiothreitol (DTT); pH 8.3), 0.025 U/µL HIV-1 RT , and different concentrations of IDA-dAMP building blocks (1mM, 500 µM, 200 µM and 100 µM). In the control reaction with the natural nucleotide, a 10 µM dATP concentration was used. The mixture was incubated at 37°C and aliquots were quenched after 5, 10, 20, 30 and 60 min. For elongation experiments, the same mixture with appropriate primer:template hybrid (see Table IV.1) was incubated at 37°C and aliquots were quenched after 15, 30, 60, 90, 120 min. In the reaction with the natural nucleotide, a 50 µM dATP concentration was used.

**Steady-state kinetics of single nucleotide incorporation**

The template T1 and the primer P1 were used. The primer and template in a 1:2 molar ratio were hybridised in a buffer containing 20 mM Tris.HCl, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.3 and used in an amount to provide 125 nM concentration of the primer in

each 20 µL reaction. A range of building block concentrations between 10 µM and 1mM for the phosphoramidate and between 0.1 µM and 10 µM for the natural building block was used. The final concentrations of primer-template complex and HIV-1 RT were 125 nM, and 0.0063 U/µL, respectively. Reaction mixtures were incubated at 37°C and aliquots were drawn at 6 different time intervals. Quenching of the reaction and polyacrylamide gel-based electrophoresis were carried out as previously described. The incorporation rates (V) were calculated based on the percentage of the extended oligonucleotide in the mixture (P+1 band). The kinetic parameters ($V_{Max}$ and $K_M$ were determined by plotting V ($nM/min^{-1}$) versus substrate concentration (µM) and fitting the data to a non-linear Michaelis-Menten regression using GraphPad Prism Software version 5.0.

### IV.4.5 Molecular modelling

Modelling of the iminodiacetate-dAMP in the HIV-1 RT active site was prepared and performed as previously described (116-118). Solvated Molecular Dynamics was used to verify the stability. The complex was solvated and simulated for 1 ns. The presence of the base pair hydrogen bonds between primer and template strand was taken as a valid indicator for the stability of the structures. Nearly all H-bonds were still present in the final dynamic structures. An average structure of the last 50 ps was generated and analyzed.

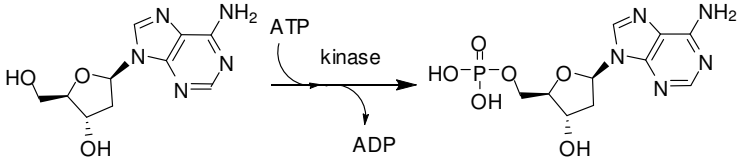
# V. Chapter five: Proposition of biosynthesis, metabolization and biodegradation of XNA building blocks for diversifying nucleic acid synthesis *in vivo*

## *V.1. Introduction*

In our efforts to diversify the synthesis of nucleic acids *in vivo*, we have identified iminodiacetic acid as a substitute for the pyrophosphate leaving group in the polymerization of DNA by HIV-1 reverse transcriptase (see chapter IV). In order to be used in artificial enzymatic DNA synthesis, alternative DNA precursors should be metabolically available and recyclable. As presented in chapter I, the deoxynucleoside triphosphates used as precursors in the natural DNA synthesis machinery are the products of successive phosphorylation reactions carried out by an assortment of phosphoryltransferases, or kinases. This succession of reactions, referred to as the 'phosphorylation cascade' (figure I.5), has drawn the attention of enzymologists over the past decades. Nucleoside kinases (NK) process nucleosides as substrates and are base-specific. Nucleotide kinases (NMK) phosphorylate nucleotides to afford nucleoside diphosphate, which are in turn phosphorylated by nucleoside diphosphate kinases (NDK). Kinases may vary in terms of amino acid sequence, tridimensional shape and or polymerization pattern (number of enzyme units in the catalytic complex and heterogeneity of this complex) depending on the organism and or nucleoside, but the mechanism of phosphoryl transfer remains universal. A functional domain accommodating the phosphate donor is found in all kinases. It is meant to accommodate polyanionic species featuring great intramolecular repulsion forces and is referred to as the 'ATP binding pocket' (125).

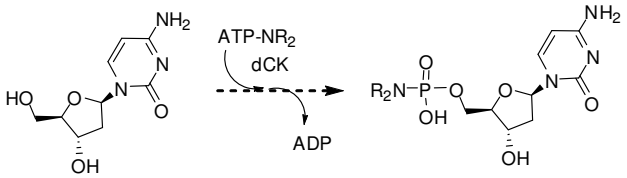
The primary phosphorylation amounts to transfering the terminal phosphate group of a 'phosphate donor' to the 5′ hydroxyl of a (deoxy)nucleoside. For instance, deoxycytidine kinase (dCK) assigns the gamma phosphate of adenosine triphosphate (ATP) to dC, dG and dA, as shown in scheme V.1. (126) While nucleoside specificity is well documented, specificity towards the phosphate donor has been less studied, ATP being more or less designated as optimal and universal donor. Depending on the kinases, dATP or UTP are also common phosphate donors. Other nucleoside triphosphates can serve for this function, albeit moderately. Interestingly, it has been recently demonstrated that CDK2, a typical wild type protein kinase, can use a series of γ-substituted ATP analogues for phosphorylation of the

$p27^{kip1}$ acceptor substrate (127), resulting in the transfer of a substituted phosphate onto a protein.




**Scheme V.1:** The first phosphorylation step affording dAMP from dA, using ATP as phosphate donor is carried out by deoxycytidine kinase.

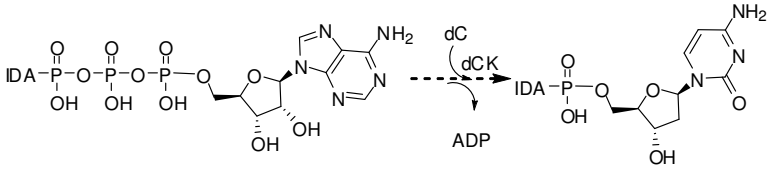
The first tridimensional crystal structure of a kinase (T4dK) was published by Teplyakov (128). Eriksson listed the structural elucidation of three additional phosphorylases in his 2002 review and Sabini et al. published the high resolution crystal structure of dCK in 2003 (129;130). A large cleft for binding of the phosphate donor seems to be a common feature of all nucleoside kinases. While the nucleoside binding is selective for A and U, the glycine-rich loop that binds to the phosphate moiety (P-loop) is a large anion pocket. The large size of the anion pocket, the flexibility of the P-loop and the induced-fit character of the binding mechanism might allow the binding of a triphosphate analogue where the gamma phosphate has been substituted with an amino-linked chemical group. Provided that the phosphoryl transfer will not be impaired by the presence of this modification, the catalysis by dCK could result in the transfer of a substituted γ-phosphate onto a chosen deoxyribonucleoside (dC, dG or dA, see scheme V.2).



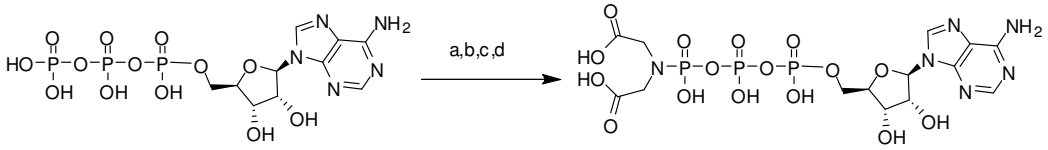

**Scheme V.2:** Potential enzymatic transfer of substituted γ-phosphoryl group onto deoxycytidine by deoxycytidine kinase (dCK).

## *V.2. Results and discussion*

Since CDK2 could use a γ-substituted triphosphate as phosphoryl donor, in this study we chose to challenge deoxycytidine kinase (dCK) with a gamma-modified ATP as a phosphate donor for the phosphorylation of one of its natural acceptor substrate deoxycytidine (scheme V.4). ATP was substituted in gamma-position with the iminodiacetic acid (IDA) following a cyclohexylcarbodiimide (CDI) approach (Scheme V.5) adapted from Hoard and Ott (131) to obtain [γ-P]-iminodiacetate adenosine triphosphate (IDA-ATP).




**Scheme V.4:** Potential phosphorylation of deoxycytidine with IDA-ATP by deoxycytidine kinase.




**Scheme V.5:** Synthetic route for iminodiacetic acid adenosine triphosphate: a) CDI, dimethylformamide, anhydrous; b) methanol, 30 min; c) iminodiacetic acid dimethyl ester; d) Sodium hydroxide, 5% in water, 15 min.

In the control phosphorylation reaction, ion exchange HPLC monitoring (Figure V.2) showed disappearance of the peak attributed to dC eluted at 11% B and of ATP at 74 % B. Disappearance of dC was simultaneous with the appearance of a small peak whose retention corresponded to that of pure dCMP (37 % B) and a bigger peak whose retention time matched that of pure ADP (57-58%). The identity of the peaks was confirmed by MS analysis of the corresponding fractions.

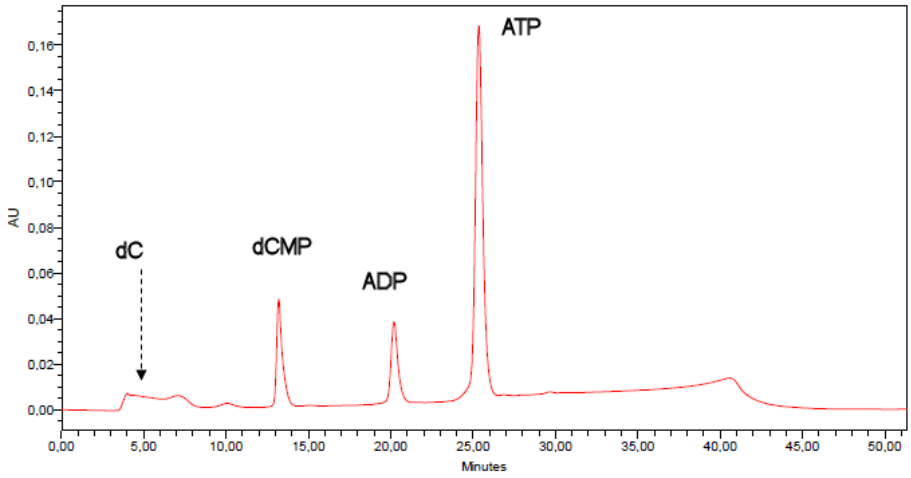



**Figure V.1:** Ion exchange HPLC profile of dC phosphorylation with ATP as phosphate donor after 16h reaction at room temperature.

Attempts to transfer the IDA-phosphate moiety to dC were monitored by ion exchange HPLC and showed that disappearance of ATP-IDA was not simultaneous with the formation of dCMP-IDA (reference peak in figure V.2). Instead, dCMP was produced, suggesting *in situ* degradation of ATP-IDA towards ATP (observed only in the presence of divalent cation) for further use as phosphate donor. The comparison of peak areas suggests that phosphorylation took place only on a small amount of dC, involving an equimolar production of ADP. The substantial amount of ADP observed after reaction can therefore not only account for the by-product of phosphorylation. This additionnal appearance of ADP can be explained by the concomitant degradation of ATP-IDA. As a matter of fact, stability of IDA-ATP in the reaction buffer was determined and ranged from 10-15 minutes at 37°C to a few hours at 25°C.

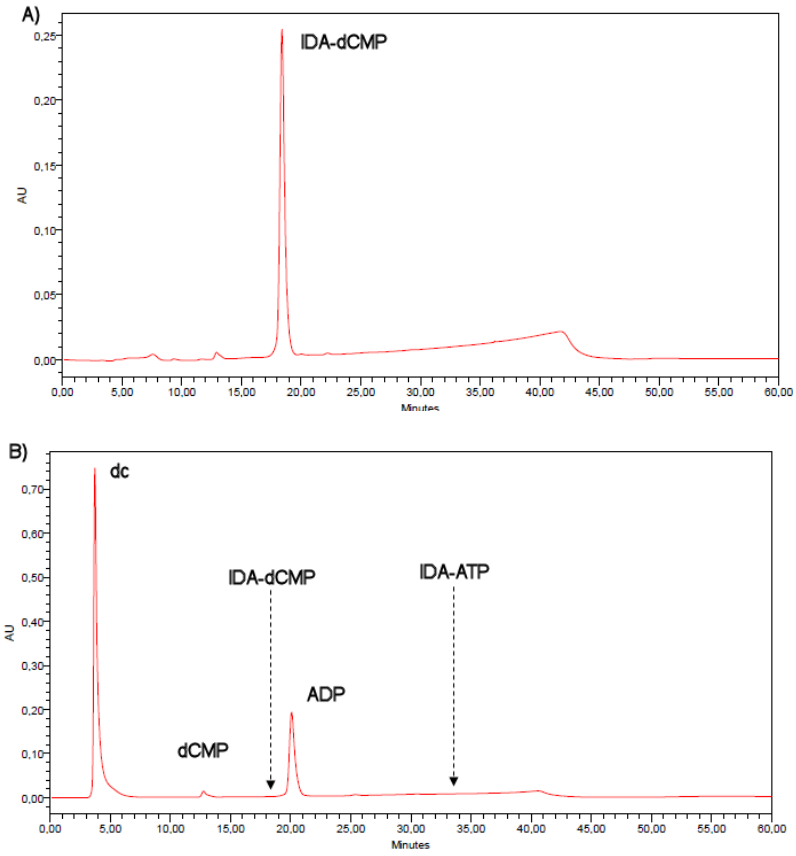



**Figure V.2:** Ion exchange HPLC profile of A) IDA-dCMP, reference peak, from chemically synthesized compound and B) Attempt of dC phosphorylation with ATP-IDA as phosphate donor: no IDA-dCMP has formed after 16h reaction at room temperature.

In addition, a recombinant human deoxycytidine kinase was used for this preliminary testing (i.e. the only commercially available deoxynucleoside kinase). Its human origin implies high substrate specificity. This restriction applies to the phosphate acceptor as well as for the phosphate donor. Testing of deoxynucleoside kinases offering broader substrate specificity, such as bacterial or archeon kinases, could allow a gamma phosphate-substituted ATP to be processed as phosphate donor towards deoxycytidine.

In the previous chapters IDA could be regarded as the 'equivalent' of two phosphate groups during the catalytic process of polymerization involving divalent cations. Therefore, after introducing IDA onto a dN triphosphate moiety, the modified substrate could be regarded as carrying five phosphate equivalents, as represented on Figure V.3. Such a compound forms a

very large anionic moiety that might not fit anymore in the anionic pocket of the enzyme's active site, impairing substrate recognition by deoxycytidine kinase.

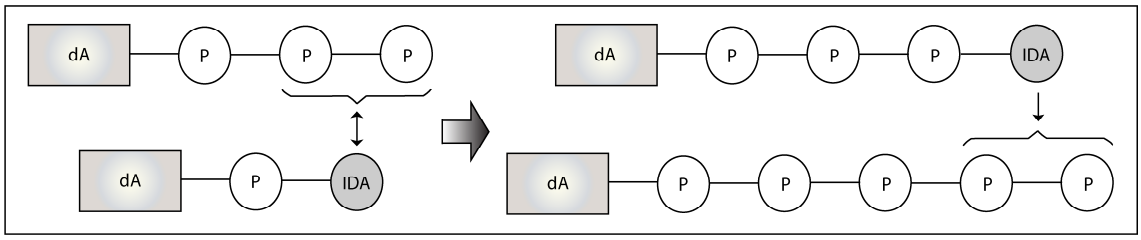



**Figure V.3:** 'Phosphate equivalence' of IDA leaving group. One group equivalent is represented as a disk.

Attempts to promote this γ-P transfer using the iminodiacetic acid is in its dimethyl ester form proved that a reduction of the number of negative charges by 2 is insufficient to promote phosphorylation, as shown on figure V.4.

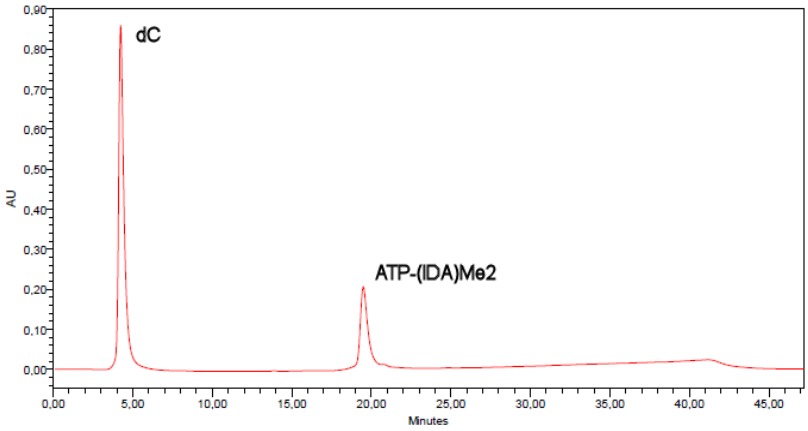



**Figure V.4:** Ion exchange HPLC profile attempting dC phosphorylation with ATP-(IDA)$Me_2$ as γ-phosphate donor after 20h reaction at room temperature: both substrates are unreacted.

All in all, the modification of the phosphate moiety of the phosphate donor species in enzymatic phosphorylation necessitates further study. It would first involve the identification of (deoxy)nucleoside kinases endowed with broader substrate specificity and then require the testing γ-P modified nucleoside triphosphates as phosphate donor substrate towards these candidate enzymes.

As introduced in Chapter I, the prerequisites for the sustainment of an orthogonal system for the synthesis and propagation of artificial informational polymers (XNA) are numerous. A proposal for such a system (with deoxyadenosine as a base) is presented in Figure V.4. As a potential alternative to pyrophosphate in the polymerization of XNA, iminodiacetic acid has

the advantage of being catabolically available. Indeed, the enzyme alanopine dehydrogenase (EC1.5.1.17) is able to catalyze the condensation reaction between an amino acid and a keto acid to form an 'opine' (132). In this reaction, nicotinamide adenine dinucleotide (NADH) could be used as reducing agent. While L-alanine and pyruvate are the preferred substrates, is has been shown that glycine and glyoxylate can also be processed, allowing the biosynthesis of iminodiacetic acid (step *a*) (133). At the other end of the polymerization loop, *in situ* metabolic degradation of the leaving group can be achieved by EDTA monooxygenase, harvested from the EDTA-degrading bacterium BNC1 (122;134). This degradation could prevent accumulation of IDA in the active site or in the 'nucleic acid factory' and could also provide a driving force for the polymerization reaction. After synthesis or recovery of IDA, it could be part of what we call an 'energetic enclave'. In this synthetic cycle, IDA replaces pyrophosphate as energetic currency. The *in situ* production of alternative building blocks is realised as presented in Chapter IV (Scheme IV.4) and involves cyclotriphosphate (a.k.a. trimetaphosphate) as a phosphate source and results in the formation of bis(iminodiacetic acid) phosphorodiamidate ($IDA_2P$) (step *b*). As a matter of fact, along with orthophosphate cyclic triphosphate is thought to have possibly served as phosphate source under pre-biotic conditions (135). The phosphoryl transfer reactions involved in steps *c)* and *d)* are not spontaneous and require to be catalytically performed. A different type of catalysis than the enzymatic one is prohibited since we aim at *in vivo* sustainment. The transfer of IDA-phosphate reaction (possibly catalyzed by a kinase) of $IDA_2P$ onto adenosine diphosphate would result in γ-P iminodiacetic acid ATP (step *c*). Released IDA is either degraded or re-enters the cycle (step *f*). Afterwards, alternative DNA building blocks carrying IDA as leaving groups would originate from the reaction between IDA-ATP and the corresponding nucleosides (step *d*), also potentially catalyzed by a kinase. At that point, the building block can enter what we call a 'genetic enclave', in which it is used by a leaving group-specific polymerase for the synthesis of the corresponding informational polymer (step *e*).

The possibility of modified phosphoryl transfer could be investigated by testing kinases from different sources or by mutating a pool of available kinases. After a suitable catalytic system is identified, the envisioned orthogonal system presented below could be assembled and tested *in vivo*.

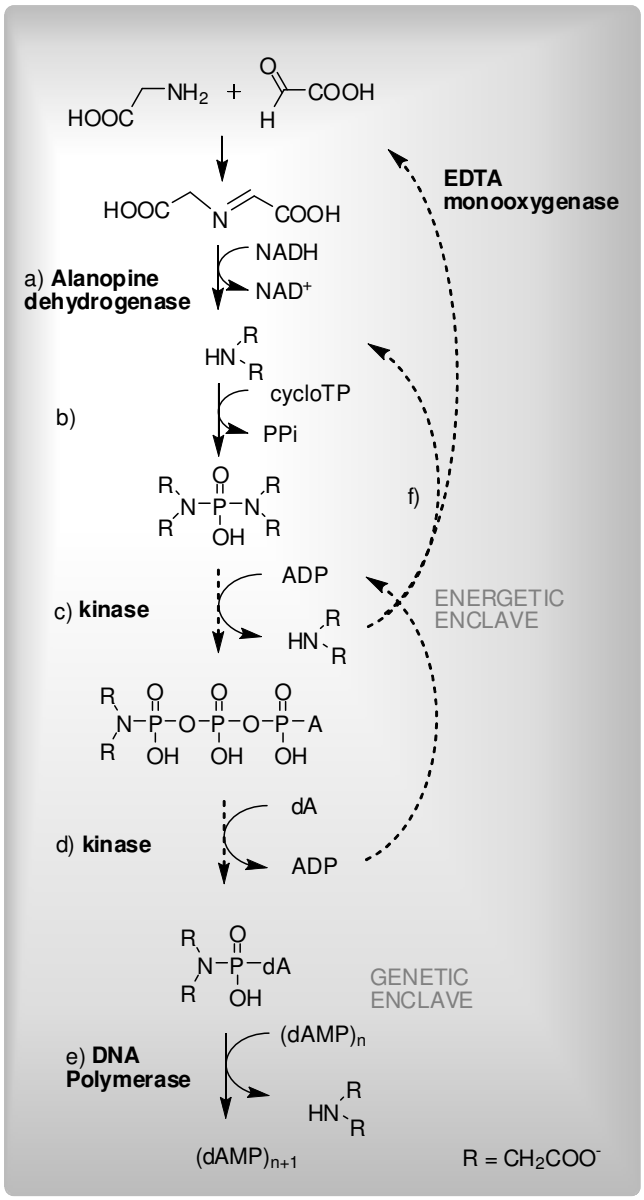



**Figure V.4:** Envisioned orthogonal system for the propagation of informational polymers using iminodiacetic acid as alternative leaving group: a) condensation of glyoxylate and glycine by an opine dehydrogenase, potentially using NADH as reducing agent; b) cyclotriphosphate opening by IDA generates alternative energetic currency; c and d) series of IDA-P transfers 'IDA-phosphorylations' yielding alternative nucleic acid precursors (exemplified for deoxyadenosine); side products IDA and ADP can re-enter the synthetic cycle and therefore promote auto-maintenance of the system; e) orthogonal synthesis of nucleic acids by specific DNA polymerase; f) possible degradation (by EDTA monooxygenase) and/or recycling (into step b) of the leaving group displaces the equilibrium in favour of nucleic acid synthesis.

## *V.4. Experimental section*

### V.4.1 General methods

NMR spectral analyses were carried out and processed as described in Chapter II. Chemical shifts are expressed in parts per million (*ppm*) by frequency. $^{1}H$ chemical shifts are referenced to an internal TMS signal (δ = 0.00 ppm), $^{31}P$ NMR chemical shifts are referenced to an external 85 % $H_3PO_4$ standard (δ = 0.00 ppm). Standard and exact mass spectra were measured as described in Chapter II. Chemicals of analytical or synthetic grade were obtained from commercial sources and were used as received unless otherwise specified (tributylamine, Acros; carbonyldiimidazole, Alfa Aesar; deoxyadenosine monophosphate: Sigma Aldrich; trimetaphosphate trisodium salt, Sigma; hexamethylphosphoramide, UCB stock). Acetonitrile HPLC Grade was purchased from Fischer Scientific. Thin Layer Chromatography was performed on Alugram® silica gel UV254 mesh 60, 0.20 mm (Macherey-Nagel). Ion exchange chromatography was performed on Source 15Q in a Tricorn$^{TM}$ column 200mmx10mm (both Pharmacia) using 2748 Dual Absorbance Detector, 600 Controller, 600 Pump with Empower 2 as processing software (all: Waters).

**DNA duplexes preparation and polyacrylamide gel-based electrophoresis** were carried out as described in Chapter II.

### V.4.2 Synthesis

***Bis(n-tributylammonium) salt of 5'-adenosine triphosphate***

Adenosine triphosphate as a disodium salt was converted to its bis(n-tributylammonium) salt by applying it as an aqueous solution onto a Dowex 50WX8-200 (Acros) ion exchange column in its protonated form and subsequently washing the column with water. The resulting free acid was dropped into a cooled solution of n-tributylamine (1.1 equiv.) in ethanol (20 mL) and further evaporation under reduced pressure afforded the product as a colourless solid.

$^{1}H$ NMR (300 MHz) $\delta$ 8.38 (s, 2H, $H_2$ & $H_8$), 5.90 (d, 1H, $^{3}J_{H1'-H2'}$ = 5.49 Hz, $H_{1'}$), 4.56 (m, 1H, $H_{3'}$), 4.38 (bs, 1H, $H_{2'}$), 4.21 (bs, 1H, $H_{4'}$), 4.11 (bs, 2H, $H_{5'}$), 2.90 (bs, 12H, C$H_2$-N), 1.45 (bs, 12H, C$H_2$-$CH_2$-N), 1.14 (d, 12H, C$H_2$-C$H_2$-$CH_2$-N), 0.73 (t, 18H, C$H_3$-$CH_2$-C$H_2$-$CH_2$-N) ppm. $^{31}P$ NMR (121 MHz) $\delta$ -10.92 (d, γ-P, $^{2}J_{\beta-\gamma}$=19.34 Hz), -11.56 (d, α-P, $^{2}J_{\alpha-\gamma}$ =19.76 Hz), -23.21 (t, β-P, $^{2}J_{\beta-\gamma} \sim {}^{2}J_{\alpha-\gamma}$ =19.59Hz) ppm.

***γ-P-imidazolide-5'-adenosine triphosphate (Im-ATP)***

Adenosine triphosphate as a bis(n-tributylammonium) salt (0.73 mmol, 460 mg) was dried overnight in vacuo over potassium oxychloride and was dissolved as a 0.1 M solution in DMF previously dried with freshly activated 0.4µm molecular sieves. Cyclohexylcarbodiimide (3.6 mmoles, 592 mg, 5 eq.) was added to the solution stirred under argon atmosphere, resulting in a clear yellow solution. The reaction was monitored by analytical ion exchange HPLC (Source 15Q running a gradient of TEAB 1M in water) coupled to ESI MS. Im-ATP was eluted at 68 % B. After 48 hours, the solution was used as such for the next synthesis step.

$^{1}$H NMR (300 MHz): $\delta$ 8.39 (s, 1H, $H_8$), 8.18 (s, 1H, $H_2$), 7.84 (s, 1H, Im), 7.21(s, 1H, Im), 6.18 (s, 1H, Im), 6.03 (d, 1H, $^{3}J_{H1'-H2'}$ = 6.03Hz, $H_{1'}$), 4.38 (m, 1H, $H_{3'}$), 4.29 (t, 1H, $H_{2'}$), 4.25 (m, 1H, $H_{4'}$), 4.03 (m, 2H, $H_{5'}$), 3.05 (m, $CH_3$-C$H_2$-$NH^+$), 1.17 (t, C$H_3$-$CH_2$-$NH^+$) ppm.

$^{31}$P NMR (121 MHz): δ –7.04 (d, P-γ, J=19.94 Hz), -11.54 (d, P-α, J=19.88 Hz), -22.49 (t, P-β, J=20.13 Hz) ppm. HRMS (ESI negative ion mode): calculated for $C_{13}H_{18}N_7O_{12}P_3$: 557.0226, found: 556.0150.

***γ-P-(iminodiacetic acid dimethyl ester) 5'-adenosine triphosphate (IDA-ATP)***

The hydrochloride salt of iminodiacetic acid dimethyl ester (5 eq., 720 mg) and n-tributylamine (5 equiv., 850 µL) were added to the previous reaction mixture and the course of the reaction was monitored by ion exchange HPLC. $Me_2$(IDA)-ATP was eluted at 50% B; Rf (iPrOH/$H_2$0/$NH_3$:5.5/3.5/1)=0.50. After 24 hours the mixture was diluted 10 times in water (Millipore) and purified by ion exchange (Source 15Q, preparative). Corresponding fractions were collected and lyophilized, affording a white fluffy solid.

$^{1}$H NMR (300 MHz): $\delta$ 8.55 (s, 1H, $H_2$), 8.26 (s, 1H, $H_8$), 6.14 (d, 1H, $^{3}J_{H1'-H2'}$ = 5.97Hz, $H_{1'}$), 4.72 (m, 1H, $H_{3'}$), 4.58 (t, 1H, $H_{2'}$), 4.39 (m, 1H, $H_{4'}$), 4.25 (m, 2H, $H_{5'}$), 3.89 (s, 2H, MeOOC-C$H_2$-N), 3.85 (s, 2H, MeOOC-C$H_2$-N), 3.67 (s, 6H, C$H_3$-OC(O)-$CH_2$-N), 3.1 (m, $CH_3$-C$H_2$-$NH^+$), 1.22 (t, C$H_3$-$CH_2$-$NH^+$) ppm. $^{31}$P NMR (121 MHz, $D_2O$): $\delta$ –3.85 (d, P-γ, $^{2}J_{\beta-\gamma}$ = 21.14 Hz), -11.51 (d, P-α, $^{2}J_{\alpha-\gamma}$ =19.03 Hz), -22.98 (t, P-β, $^{2}J_{\beta-\gamma} \sim {}^{2}J_{\alpha-\gamma)}$ =20.13 Hz) ppm. HRMS (ESI negative ion mode): calculated for $C_{16}H_{25}N_6O_{16}P_3$: 650.0540, found: 649.0471.

***γ-P-(iminodiacetic acid) 5'-adenosine triphosphate***

The dimethyl ester of iminodiacetic acid ATP was dissolved in a 5% NaOH solution and left to react at room temperature. The course of the reaction was monitored by analytical ion exchange HPLC and TLC (iPrOH:$H_2O$:$NH_3$ 5.5:3.5:1). Complete conversion to the free dicarboxylic acid (which was eluted at 82 % TEAB 1M) was observed after 30 min.; Rf

(iPrOH/$H_20$/$NH_3$:5.5/3.5/1)=0.21. The reaction mixture was purified by preparative ion exchange HPLC and the corresponding fractions were collected and lyophilized, affording a white fluffy solid.

$^1$H NMR (300 MHz): $\delta$ 8.46 (s, 1H, $H_8$), 8.17 (s, 1H, $H_2$), 6.05 (d, 1H, $^3J_{H1'-H2'}$ = 4.80Hz, $H_{1'}$), 4.55 (m, 1H, $H_{3'}$), 4.28 (t, 1H, $H_{2'}$), 4.18 (m, 1H, $H_{4'}$), 4.11 (m, 2H, $H_{5'}$), 3.63 (s, 2H, HOOC-C$H_2$-N), 3.60 (s, 2H, HOOC-C$H_2$-N), 3.10 (m, $CH_3$-C$H_2$-$NH^+$), 1.18 (t, C$H_3$-$CH_2$-$NH^+$) ppm.

$^{31}$P NMR (121 MHz): $\delta$ –3.34 (d, P-γ), -11.46 (d, P-α, $^2J_{\alpha-\gamma}$ =18.9Hz), -22.79 (t, P-β) ppm.

HRMS (ESI negative ion mode): calculated for $C_{14}H_{21}N_6O_{16}P_3$: 622.0227, found: 621.0168.

## V.4.3 Enzymatic assays

***Transfer of γ-P-IDA onto 5'-OH of deoxyadenosine***

A solution of 0.5 mM ATP or IDA-ATP, 0.5 mM dCMP and 0.35 µM deoxycytidine kinase (Biaffin GmbH), in kinase buffer (50mM Tris.HCl, 100mM KCL, 5mM $MgCl_2$, dithiothreitol 1 mM), for a total reaction volume of 200 µL, was left at room temperature for 16 hours. The reaction mixture was then purified on an ion exchange column (Source 15Q) eluting with a gradient of TEAB 1M in water (0 to 100 % in 35 min).

# VI. Conclusions and perspectives

The polymerization of nucleotides into nucleic acids is catalyzed by polymerases and is the supporting process for the survival and reproduction of all living organisms. Four deoxynucleoside triphosphates (dATP, dCTP, dGTP and dTTP) are the traditional precursors of deoxyribonucleic acid (DNA). In recent decades numerous analogues of these precursors have been designed and synthesized in order to study the mechanism of their enzymatic polymerization. For this purpose, modifications were synthetically introduced on different parts of the molecule, mainly on the nucleobase or the sugar-phosphate backbone. Polymerase recognition of the corresponding deoxynucleoside triphosphate analogues can be characterized and inform us about the importance of enzyme-substrate binding interactions. In this context, modifications of pyrophosphate were not explored until very recently, since it is not part of the resulting biopolymer *per se*. Nevertheless, initial binding of the dNTP to active site residues play a key role on polymerization efficiency and replication fidelity. Therefore, enzyme recognition of dNTPs analogues where pyrophosphate is replaced by various leaving group candidates can instruct us about the role of pyrophosphate in the polymerization mechanism and its influence on different stages of this biochemical event.

Deoxyadenosine triphosphate analogues where the pyrophosphate moiety is replaced by L-aspartic acid and few other amino acids were previously shown to act as substrates for DNA polymerization catalyzed by promiscuous polymerases such as HIV-1 reverse transcriptase (RT) or Vent exo$^-$ polymerase. It was suggested that the functional moieties of these alternative leaving groups (carboxylic acid and, or imidazole ring) could act as mimics of pyrophosphate in the enzyme's active site. The aim of this thesis was to identify new leaving group candidates that are able to replace pyrophosphate during nucleic acid synthesis. For this purpose, structures similar to L-Asp and L-His in shape, size and coordination capacity are explored.

In **chapter two** of this thesis it was demonstrated that a variety of other chemical groups with potential coordination capacity could also be processed as leaving groups during HIV-1 RT catalyzed DNA synthesis. Coordination capacity, relative position of the potential coordinating groups, as well as the phosphoramidate or phosphodiester nature of the linkage between leaving group and nucleotide, seemed to affect the incorporation outcome.

Results presented in **chapter three** completed the evaluation of the influence of the linking atom in relation to the efficiency of nucleotide incorporation. For this purpose, phosphodiester analogues of the best performing amino acid dAMP phosphoramidates were synthesized. Regrettably, poor stability of malic acid dAMP (phosphodiester analogue of L-Asp-dAMP) in polymerization media prevented any enzymatic tests. L-β-imidazole lactic acid dAMP (L-ILA dAMP, phosphodiester analogue of L-His-dAMP) proved to be just stable enough to allow its evaluation as a substrate for reverse transcriptase. As expected, L-ILA was able to mimic the pyrophosphate moiety as leaving group in the DNA polymerization reaction. Replacement of a phosphoramidate linkage by a phosphodiester linkage resulted in greater substrate affinity, higher maximum velocity and fairly better elongation capacity. Unfortunately full-length primer elongation opposite a template with a seven base overhang could not be obtained with either of the dATP analogues. Many reasons could explain premature DNA synthesis termination, such as the absence of energy, usually provided by pyrophosphate hydrolysis, and required by the enzyme for further conformational changes. DNA elongation could also be hampered by uncorrect binding of the substrate in the active site (leading to substrate inhibition of the enzyme), or strong binding of the product in or out the active site (leading to product inhibition). Probably due to instability of the phosphorothioate analogues of L-Asp-, β-Ala- and GA-dAMP, the compounds could not be isolated. The influence of replacing the linking oxygen by the other possible isostere sulphur could therefore not be investigated.

As detailed **in chapter four**, great incorporation and elongation results were obtained when using iminodiacetic acid dAMP as substrate. When this analogue was used as the substrate, the full-length elongation product was partially obtained. Moreover, supplementary elongation results also demonstrated the positive influence of a longer template overhang for the recognition and incorporation of several successive units carrying a modified leaving group, thus validating the importance of interactions between finger subdomain residues and the template's upstream nucleotides in the HIV-1 RT polymerization process. The calculated model of the ground state interactions for accommodation of IDA-dAMP in the active site of reverse transcriptase showed that similarly to pyrophosphate, iminodiacetic acid seems to be bound in a network of catalytical residues of the enzyme with the help of two catalytical metal ions. This molecular model supports the idea that IDA probably mimics pyrophosphate in this mechanism by coordinating essential features in the active site (catalytic ions and residues). Overall, comparison of the results obtained with our series of analogues points out that IDA-

dAMP is the best analogue and hence IDA the best alternative leaving group for pyrophosphate in the enzymatic polymerization by HIV-1 reverse transcriptase.

The synthetic work reported in this thesis is centred on phosphorous coupling chemistry. Many methods are available and allow the formation of P-O and P-N bonds, ranging from some requiring rigorous anhydrous conditions to some more user- and environment-friendly reactions carried out in water solution. Phosphorous coupling chemistry has enormously developed since the importance of ATP and that of the phosphodiester bond in the life cycle have been uncovered. Phosphorous reactivity remains a matter of investigation, since polyphosphates are now thought to be the molecular source of energy in prebiotic conditions on earth. The presence of one or more phosphate groups in test compounds confers them a very high polarity. Their purification requires in turn very specific and polar methods like ion exchange chromatography and the use of trialkylammonium salts to promote dissolution. Unfortunately for the chemist, this greatly reduces the yield of coupling reactions since much product is lost over numerous steps of HPLC purification and subsequent lyophilisation. Therefore, the search for innovative synthetic and purification strategies deserves to be encouraged.

Studying a chemical group in a biochemical environment and the effects of its modification on a biochemical reaction is not a simple task. Biochemical actors such as proteins endowed with catalytic activity add a tremendous amount of uncertainty to *in vitro* experiments, whose results should therefore be interpreted with uttermost circumspection. On a practical point of view, so far only the crystallographic resolution of static structures is possible, which limits the observation to discrete steps of the process and complicates the examination of transition steps. In addition, working with enzymes greatly decreases the reproducibility of experimental results. As an illustration, standard errors of a remarkably high 30 % are considered as acceptable by standard textbooks (136). Furthermore, inter-batch variations in specific activity (activity per mg enzyme) and *a fortiori* changes in enzyme manufacturer, shipping and storing conditions can interfere considerably with experimental reproducibility. Nevertheless, experimental results of *in vitro* enzymatic reactions provide valuable insights on their mechanism, helping the medicinal chemist to better understand structure activity relationships of potential tools for biotechnology or drug therapies. Conclusions drawn in the three first studies of this thesis are consistent with each other and converge towards various postulations. First, it is possible to **replace pyrophosphate** in the polymerization of DNA with **various alternative leaving groups**, albeit, until now, at the expense of incorporation

and elongation efficiency. Second, intrinsic physicochemical properties (size, coordinating capacity, spatial arrangement, and polarisability of the phosphorous linkage; pKa and stability of the released species) of the pyrophosphate mimic can differ drastically from that of the natural and so far optimal leaving group. Therefore this replacement is likely to have **direct effects on the chemical step** (equilibrium displacement, rate, yield and geometry of successive states). Interestingly, better incorporation efficiency does not correlate with a enhanced elongation capacity. Third, altering the phosphorous linkage and or the structure of the leaving group might also **affect previous and following steps** of the overall biochemical event.

Among other characteristics of the leaving group, its **affinity** towards the duplex DNA-enzyme complex during the second binding step is thought to influence substrate selectivity and polymerization kinetics (137). In this regard, modelling experiments provide indications on coordinating possibilities, but crystallization of HIV-1 RT liganded with dsDNA and IDA-ddNMP remains the best method to study ground state interactions between alternative leaving group and active site's residues.

Downstream in the polymerization process, released pyrophosphate is suspected to have an important role in catalytic ion removal, DNA duplex translocation to the following incorporation site (108) or overall conformational changes of successive enzymatic complexes (92). The **biochemical fate** of a modified leaving group will certainly differ from that of pyrophosphate depending on its coordinating capacity and interaction with its environment. For instance, an explanation for premature DNA synthesis termination witnessed in chapter III could be that released imidazole lactic acid or L-histidine binds to reverse transcriptase, thereby inhibiting the enzyme. Since rates reported from gel bands in percent extension per minute are difficult to interpret precisely and therefore prohibit mechanistic insights. Doubtless determination of such a situation requires using analytical tools involving fluorescence detection of product formation and/or consumption (138). As proposed by Götte, the absence of released PPi and corresponding free enthropy 'cellular energy' in the mechanism might account for the loss of processivity and or efficiency of reverse transcriptase observed in elongation experiments after incorporation of a few nucleotides (92). As stressed out by Johnson in a recent review (138), the role of pyrophosphate in the polymerization mechanism remains an area of active investigation. By reviewing the kinetics and the chemical mechanism of DNA polymerases, Johnson reminds us that the (free) energy source required for enzyme isomerisation remains undetermined. In this

regard, his team judiciously developed a sensitive technique allowing real time measurement of PPi release that could be used to further investigate the relationship between PPi release and DNA polymerization (139). Additional fundamental research should be done to gain more information on the role of other intrinsic properties of the leaving group as well as short and long term effect of its modifications on the polymerization mechanism. In particular, efforts should be put towards developing novel analytical tools allowing real time visualisation of biochemical complexes.

The study of DNA polymerization helps understanding a key step of viral replication. New insights on its mechanism could help future design of powerful antiviral compounds. Nucleoside like reverse transcriptase inhibitors (NRTIs) are broadly used in antiviral drug regimens, but the development of drug resistant strains leads to a rapid decrease in therapeutic efficacy. There is therefore a constantly renewed need for novel antiviral compounds, and it is our strong belief that NRTI analogues containing a pyrophosphate mimic such as iminodiacetic acid could be used efficiently as therapeutic compounds. Conjugating an RT inhibitor which is not substrate for cellular kinases, with a pyrophosphate mimic *via* a phosphoramidate linkage, could allow by-passing of the phosphorylation cascade and promote the delivery of an already active dNTP analogue into targeted cells. McGuigan and co-workers have been addressing the search for new antiviral compounds with an approach involving a lipophilic group and an amino acid phosphoramidate, but unfortunately no studies were carried out with potential pyrophosphate mimics so far. In this original approach, the alternative leaving group would serve both goals of masking the nucleoside monophosphate and by-passing the two remaining phosphorylation reactions for efficient inhibition of viral DNA replication.

Meanwhile, the timeframe and efficacy constraints of the project for orthogonalization of nucleic acid synthesis demands that a leaving group be chosen out of the series of candidates. From our studies, which are limited to deoxyadenosine triphosphate analogues, the most promising alternative to pyrophosphate is undoubtedly iminodiacetic acid. It is superior to the other analogues by its stability and substrate efficiency for incorporation and elongation. Discrepancies in the results obtained by colleagues with T, C and G analogues showed that the efficiency of IDA at mimicking pyrophosphate strongly depends on the nucleobases, which prevents for now an accurate evaluation of polymerization selectivity and specificity. The reason for this nucleobase dependence is still a matter of investigation. Further evaluation

of IDA as the orthogonal leaving group candidate would also require challenging the polymerases for much longer elongation. In the search of better leaving group candidates, analogous dicarboxylic acids are being evaluated in the laboratory of medicinal chemistry. Among other tracks, the influence of the length of the alkyl chain connecting the nitrogen atom to the carboxylic acid functions is being investigated, since a dicarboxybutyl phosphoramidate analogue of dCTP showed surprising incorporation capacities (140). Besides, the possibility of replacing the moieties with coordinating potential by phosphonic acids is also being investigated. The near future of the project for DNA synthesis orthogonalization now depends on the ability of polymerases to evolve in such a way that they favour the modified leaving group over the natural one, and eventually to segregate fully modified XNA precursors (unnatural base pair linked to a modified sugar –phosphate backbone activated by a pyrophosphate mimic) from natural DNA precursors.

Further development of an orthogonal genetic system requires that the leaving group candidate is metabolically available and degradable. Attempts for the chemical synthesis of IDA-dAMP from biological precursors were reported at the end of chapter four and the need for enzymatic catalysis was suggested. In **chapter five** we aimed at introducing an independent metabolic route for the biosynthesis of *in vivo* precursors of artificial nucleic acids. The search for broad substrate kinases able to transfer a phosphoro-iminodiacetic acid group onto a deoxynucleoside could be initiated by screening the numerous phosphoryltransferases that can be obtained from bacterial sources. Potential metabolization of DNA precursors carrying the selected alternative leaving group seems therefore reasonable. The possibility of iminodiacetic acid degradation (avoiding *in situ* accumulation and potentially providing a driving force for polymerization) already exists since EDTA monooxygenase has been identified as a bacterial enzyme endowed with appropriate activities and could be successfully cloned. Adaptation of existing enzymes to exogenous nucleic acid precursors is now in the hand of evolutionary biologists and, in this context we are expecting much of innovative methods for bacterial growth on modified substrates. The efforts developed nowadays in this direction, as well as the rapid expansion of synthetic biology augur well for the feasibility of an orthogonal XNA synthesis machinery.

Reference List

Reference List

1. Watson,J.D. and Crick,F.H. (1953) The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol*, **18**, 123-131.

2. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag.

3. Kornberg,A. and Baker,T. (1992) *DNA replication*. University Science Books.

4. Kool,E.T. (2002) Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.*, **71**, 191-219.

5. Kornberg,A., Lehman,I.R., Bessman,M.J. and Simms,E.S. (1956) Enzymic Synthesis of Deoxyribonucleic Acid. *Biochimica et Biophysica Acta*, **21**, 197-198.

6. Lamers,M.H. and O'Donnell,M. (2008) A consensus view of DNA binding by the C family of replicative DNA polymerases. *Proc. Natl. Acad. Sci. U. S. A*, **105**, 20565-20566.

7. Battula,N. and Loeb,L.A. (1976) Fidelity of DNA-replication - Lack of exodeoxyribonuclease activity and error-correcting function in avian-myeloblastosis virus DNA-polymerase. *Journal of Biological Chemistry*, **251**, 982-986.

8. Roberts,J.D., Preston,B.D., Johnston,L.A., Soni,A., Loeb,L.A. and Kunkel,T.A. (1989) Fidelity of 2 retroviral reverse transcriptases during DNA-dependent DNA-synthesis *in vitro*. *Molecular and Cellular Biology*, **9**, 469-476.

9. Bakhanashvili,M. and Hizi,A. (1992) Fidelity of the reverse-transcriptase of human-immunodeficiency-virus type-2. *Febs Letters*, **306**, 151-156.

10. Braithwaite,D.K. and Ito,J. (1993) Compilation, alignment, and phylogenetic-relationships of DNA-polymerases. *Nucleic Acids Research*, **21**, 787-802.

11. Negroni,M. and Buc,H. (2001) Retroviral recombination: what drives the switch? *Nat. Rev. Mol. Cell Biol*, **2**, 151-155.

12. Steitz,T.A. (1999) DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.*, **274**, 17395-17398.

13. Aravind,L. and Koonin,E.V. (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Research*, **26**, 3746-3752.

14. Doublie,S. and Ellenberger,T. (1998) The mechanism of action of T7 DNA polymerase. *Curr. Opin. Struct. Biol*, **8**, 704-712.

15. Pelletier,H., Sawaya,M.R., Wolfle,W., Wilson,S.H. and Kraut,J. (1996) Crystal structures of human DNA polymerase beta complexed with DNA: implications for catalytic mechanism, processivity, and fidelity. *Biochemistry*, **35**, 12742-12761.

16. Sawaya,M.R., Pelletier,H., Kumar,A., Wilson,S.H. and Kraut,J. (1994) Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism. *Science*, **264**, 1930-1935.

17. Bebenek,K., Abbotts,J., Roberts,J.D., Wilson,S.H. and Kunkel,T.A. (1989) Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse-transcriptase. *Journal of Biological Chemistry*, **264**, 16948-16956.

18. Huang,H., Chopra,R., Verdine,G.L. and Harrison,S.C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669-1675.

19. Reese,C.B. (2005) Oligo- and poly-nucleotides: 50 years of chemical synthesis. *Organic & Biomolecular Chemistry*, **3**, 3851-3868.

20. Matteucci,M.D. and Caruthers,M.H. (1981) Synthesis of deoxyoligonucleotides on a polymer support. *J.Am.Chem.Soc.*, **103**, 3185-3191.

21. Hayakawa,Y. (1986) Nonaqueous oxidation of nucleoside phosphites to the phosphates. *Tetrahedron Letters*, **27**, 4191-4194.

22. Uzagare M.C. (2003) NBS-DMSO as a nonaqueous nonbasic oxidation reagent for the synthesis of oligonucleotides. *Bioorg. & Med. Chem. Lett.*, **13**, 3537-3540.

23. Yoshikawa,M., Kato,T. and Takenish.T (1967) A novel method for phosphorylation of nucleosides to 5'-nucleotides. *Tetrahedron Letters*, 5065-5068.

24. Ludwig,J. (1981) A new route to nucleoside 5'-triphosphates. *Acta Biochim. Biophys. Acad. Sci. Hung.*, **16**, 131-133.

25. Smith,M. and Khorana,H.G. (1958) Nucleoside polyphosphates .6. An improved and general method for the synthesis of ribonucleoside and deoxyribonucleoside 5'-triphosphates. *J. Am. Chem. Soc.*, **80**, 1141-1145.

26. Abraham,T.W. (1996) Synthesis and biological activity of aromatic amino acid phosphoramidates of 5-fluoro-2'-deoxyuridine and 1-beta-arabinofuranosylcytosine: evidence of phosphoramidase activity. *J. Med. Chem.*, **39**, 4569-4575.

27. Cramer,F. and Neunhoeffer,H. (1962) Zur chemie der energiereichen phosphate .15. Reaktionen von adenosin-5'-phosphorsaure-imidazolid-eine neue synthese von adenosindiphosphat und flavin-adenin-dinucleotid. *Chemische Berichte-Recueil*, **95**, 1664-1669.

28. Moffat,J.G. and Khorana,H.G. (1961) Nucleoside polyphosphates. X. The synthesis and some reactions of nucleoside-5'-phosphoromorpholidates and related compounds. *J. Med. Chem.*, **83**, 649-658.

29. Lohrmann,R. and Orgel,L.E. (1978) Preferential formation of (2'-5')-linked internucleotide bonds in non-enzymatic reactions. *Tetrahedron*, **34**, 853-855.

30. Sawai,H. and Orgel,L.E. (1975) Oligonucleotide synthesis catalyzed by the $Zn^{2+}$ ion. *J. Am. Chem. Soc.*, **97**, 3532-3533.

31. Mukaiyama,T. and Hashimoto,M. (1971) Phosphorylation by oxidation-reduction condensation. Preparation of active phosphorylating reagents. *Bull. Chem. Soc. Jap.*, **44**, 2284-2284.

32. Kers,A., Stawinski,J. and Kraszewski,A. (1995) Studies on aryl H-phosphonates; part 2 : A general method for the preparation of alkyl H-phosphonate monoesters. *Synthesis*, 427-430.

33. Lin,C., Fu,H., Tu,G. and Zhao,Y. (2004) Synthesis of AZT/d4T boranophosphates as anti-HIV prodrug candidates. *Synthesis*, **4**, 0509-0516.

34. Atherton,F.R. and Todd,A.R. (1945) Studies on phosphorylation. Part II. The reaction of dialkyl phosphites with polyhalogen compounds in presence of bases. A new method for the phosphorylation of amines. *J. Chem. Soc.*, 660-663.

35. Garegg,P.J., Stawinski,J. and Stromberg,R. (1985) Formation of internucleotidic bonds via phosphonate intermediates. *Chem. Scripta*, **25**, 280-282.

36. Garegg,P.J. and Stromberg,R. (1987) Nucleoside H-phosphonates. V. The mechanism of hydrogenphosphonate diester formation using acylchlorides as coupling agents in oligonucleotide. *Nucleos. & Nucleot.*, **6**, 655-662.

37. Garegg,P.J., Stromberg,R. and Stawinski,J. (1987) Studies on the oxidation of nucleoside phosphonates. *Nucleos. & Nucleot.*, **6**, 429-432.

38. Wada,T. (1998) A convenient method for phosphorylation involving a facile oxidation of H-phosphonate monoesters via Bis(trimethylsilyl) phosphites. *Tetrahedron Letters*, **39**, 7123-7126.

39. Michalski,J. and Dabkowski,W. (2004) State of the art. Chemical synthesis of biophosphates and their analogues via PIII derivatives. *Top. Curr. Chem.*, 93-144.

40. Miller,P.S., Agris,C.H., Aurelian,L., Blake,K.R., Murakami,A., Reddy,M.P., Spitz,S.A. and Ts'o,P.O. (1985) Control of ribonucleic acid function by oligonucleoside methylphosphonates. *Biochimie*, **67**, 769-776.

41. Van Aerschot,A. (2006) Oligonucleotides as antivirals: dream or realistic perspective? *Antiviral Res.*, **71**, 307-316.

42. Nielsen,P.E. (2004) PNA technology. *Molecular Biotechnology*, **26**, 233-248.

43. Nauwelaerts,K., Lescrinier,E., Sclep,G. and Herdewijn,P. (2005) Cyclohexenyl nucleic acids: conformationally flexible oligonucleotides. *Nucleic Acids Res.*, **33**, 2452-2463.

44. Famulok,M. (1999) Oligonucleotide aptamers that recognize small molecules. *Curr. Opin. Struct. Biol.*, **9**, 324-329.

45. Agrawal,S. and Zhao,Q.Y. (1998) Antisense therapeutics. *Current Opinion in Chemical Biology*, **2**, 519-528.

46. Mullis,K., Faloona,F., Scharf,S., Saiki,R., Horn,G. and Erlich,H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol*, **51 Pt 1**, 263-273.

47. Maxam,A.M. and Gilbert,W. (1992) A new method for sequencing DNA. 1977. *Biotechnology*, **24**, 99-103.

48. Sanger,F. and Coulson,A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol*, **94**, 441-448.

49. Hall,N. (2007) Advanced sequencing technologies and their wider impact in microbiology. *J. Exp. Biol*, **210**, 1518-1525.

50. Schuster,S.C. (2008) Next-generation sequencing transforms today's biology. *Nat. Methods*, **5**, 16-18.

51. Niemeyer,C.M. (1997) DNA as a material for nanotechnology. *Angewandte Chemie-International Edition in English*, **36**, 585-587.

52. Rothemund,P.W. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature*, **440**, 297-302.

53. Wengel,J. (2004) Nucleic acid nanotechnology-towards Angstrom-scale engineering. *Org. Biomol. Chem.*, **2**, 277-280.

54. Ye,X., Al-Babili,S., Kloti,A., Zhang,J., Lucca,P., Beyer,P. and Potrykus,I. (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*, **287**, 303-305.

55. Shinmyo,A. and Kato,K. (2010) Molecular farming: production of drugs and vaccines in higher plants. *J. Antibiot. (Tokyo)*, **63**, 431-433.

56. Herdewijn,P. and Marlière,P. (2009) Towards safe genetically modified organisms through the chemical diversification of nucleic acids. *Chemistry & Biodiversity*, **6**, 791-807.

57. El Safadi,Y., Vivet-Boudou,V. and Marquet,R. (2007) HIV-1 reverse transcriptase inhibitors. *Appl. Microbiol. Biotechnol.*, **75**, 723-737.

58. Kakuda,T.N. (2000) Pharmacology of nucleoside and nucleotide reverse transcriptase inhibitor-induced mitochondrial toxicity. *Clin. Ther.*, **22**, 685-708.

59. Kempeneers,V., Renders,M., Froeyen,M. and Herdewijn,P. (2005) Investigation of the DNA-dependent cyclohexenyl nucleic acid polymerization and the cyclohexenyl nucleic acid-dependent DNA polymerization. *Nucleic Acids Res.*, **33**, 3828-3836.

60. Pochet,S., Kaminski,A., Van Aerschot,A., Herdewijn,P. and Marlière,P. (2003) Replication of hexitol oligonucleotides as a prelude to the propagation of a third type of nucleic acid in vivo. *C. R. Biologies*, **326**, 1175-1184.

61. Tae,E.L., Wu,Y., Xia,G., Schultz,P.G. and Romesberg,F.E. (2001) Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs. *J. Am. Chem. Soc.*, **123**, 7439-7440.

62. Malyshev,D.A., Seo,Y.J., Ordoukhanian,P. and Romesberg,F.E. (2009) PCR with an expanded genetic alphabet. *J. Am. Chem. Soc.*, **131**, 14620-14621.

63. Seo,Y.J., Matsuda,S. and Romesberg,F.E. (2009) Transcription of an expanded genetic alphabet. *J. Am. Chem. Soc.*, **131**, 5046-5047.

64. Neumann,H., Wang,K.H., Davis,L., Garcia-Alai,M. and Chin,J.W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, **464**, 441-444.

65. Kuchner,O. and Arnold,F.H. (1997) Directed evolution of enzyme catalysts. *Trends Biotechnol.*, **15**, 523-530.

66. Fernandez-Gacio,A., Uguen,M. and Fastrez,J. (2003) Phage display as a tool for the directed evolution of enzymes. *Trends in Biotechnology*, **21**, 408-414.

67. Ghadessy,F.J. and Holliger,P. (2007) Compartmentalized self-replication: a novel method for the directed evolution of polymerases and other enzymes. *Methods Mol. Biol.*, **352**, 237-248.

68. Ghadessy,F.J., Ramsay,N., Boudsocq,F., Loakes,D., Brown,A., Iwai,S., Vaisman,A., Woodgate,R. and Holliger,P. (2004) Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol.*, **22**, 755-759.

69. Ghadessy,F.J., Ong,J.L. and Holliger,P. (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci U. S. A*, **98**, 4552-4557.

70. Sucato,C.A., Upton,T.G., Kashemirov,B.A., Batra,V.K., Martinek,V., Xiang,Y., Beard,W.A., Pedersen,L.C., Wilson,S.H., McKenna,C.E. *et al.* (2007) Modifying the beta,gamma leaving-group bridging oxygen alters nucleotide incorporation efficiency, fidelity, and the catalytic mechanism of DNA polymerase beta. *Biochemistry*, **46**, 461-471.

71. Sucato,C.A., Upton,T.G., Kashemirov,B.A., Osuna,J., Oertell,K., Beard,W.A., Wilson,S.H., Florian,J., Warshel,A., McKenna,C.E. *et al.* (2008) DNA polymerase beta fidelity: Halomethylene-modified leaving groups in pre-steady-state kinetic analysis reveal differences at the chemical transition state. *Biochemistry*, **47**, 870-879.

72. Nogrady,T. and Weaver,D.F. (2005) Drug molecules: Structures and properties. *Medicinal Chemistry: A Molecular and Biochemical Approach*. Oxford University Press Inc., New York, pp. 9-66.

73. Helgstrand,E., Eriksson,B., Johansson,N.G., Lannero,B., Larsson,A., Misiorny,A., Noren,J.O., Sjoberg,B., Stenberg,K., Stening,G. *et al.* (1978) Trisodium phosphonoformate, a new anti-viral compound. *Science*, **201**, 819-821.

74. Hutchinson,D.W. (1985) Metal chelators as potential antiviral agents. *Antiviral Research*, **5**, 193-205.

75. Drontle,D.P. and Wagner,C.R. (2004) Designing a pronucleotide stratagem: lessons from amino acid phosphoramidates of anticancer and antiviral pyrimidines. *Mini-rev. in Med. Chem.*, **4**, 409-419.

76. Van Rompay,A.R., Johansson,M. and Karlsson,A. (2003) Substrate specificity and phosphorylation of antiviral and anticancer nucleoside analogues by human deoxyribonucleoside kinases and ribonucleoside kinases. *Pharmacology & Therapeutics*, **100**, 119-139.

77. Wagner,C.R., Iyer,V.V. and McIntee,E.J. Pronucleotides: Toward the *in-vivo* delivery of antiviral and anticancer nucleotides. *Medicinal Research Reviews* , 417-451. 2000. Wiley & Sons.

78. De Clercq,E. (2002) Strategies in the design of antiviral drugs. *Nature Reviews Drug Discovery*, **1**, 13-25.

79. McGuigan,C., Bellevergue,P., Sheeka,H., Mahmood,N. and Hay,A.J. (1994) Certain phosphoramidate derivatives of dideoxy uridine (ddU) are active against HIV and successfully by-pass thymidine kinase. *FEBS Lett.*, **351**, 11-14.

80. Meier C. (1998) Pro-Nucleotides - Recent advances in the design of efficient tools for the delivery of biologically active nucleoside monophosphates. *Synlett*, 233-242.

81. Adelfinskaya,O. and Herdewijn,P. (2007) Amino acid phosphoramidate nucleotides as alternative substrates for HIV-1 reverse transcriptase. *Angewandte Chemie-International Edition*, **46**, 4356-4358.

82. Adelfinskaya,O., Terrazas,M., Froeyen,M., Marliere,P., Nauwelaerts,K. and Herdewijn,P. (2007) Polymerase-catalyzed synthesis of DNA from phosphoramidate conjugates of deoxynucleotides and amino acids. *Nucleic Acids Research*, **35**, 5060-5072.

83. Terrazas,M., Marlière,P. and Herdewijn,P. (2008) Enzymatically catalyzed DNA synthesis using L-Asp-dGMP, L-Asp-dCMP, and L-Asp-dTMP. *Chemistry & Biodiversity*, **5**, 31-39.

84. Garforth,S.J., Parniak,M.A. and Prasad,V.R. (2008) Utilization of a deoxynucleoside diphosphate substrate by HIV reverse transcriptase. *PLos One*, **3**, e2074.

85. Sawai,H. (1990) Divalent metal ion-mediated phosphodiester bond formation from adenosine-5'-phosphorimidazolide with glycolic acid or lactic acid in aqueous pyridine. A nonenzymatic model reaction for nucleotidyl transfer. *Bull. Chem. Soc. Jap.*, **63**, 692-696.

86. Mukaiyama,T. and Hashimoto,M. (1972) Synthesis of oligothymidylates and nucleoside cyclic phosphates by oxidation-reduction condensation. *J. Am. Chem. Soc.*, **94**, 8528-8532.

87. Boosalis,M.S., Petruska,J. and Goodman,M.F. (1987) Dna-polymerase insertion fidelity - Gel assay for site-specific kinetics. *Journal of Biological Chemistry*, **262**, 14689-14696.

88. Creighton,S., Bloom,L.B. and Goodman,M.F. (1995) Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Meth. Enzymol.*, **262**, 232-256.

89. Castro,C., Smidansky,E.D., Arnold,J.J., Maksimchuk,K.R., Moustafa,I., Uchida,A., Gotte,M., Konigsberg,W. and Cameron,C.E. (2009) Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nature Structural & Molecular Biology*, **16**, 212-218.

90. Ninio,J. and Orgel,L.E. (1978) Heteropolynucleotides as templates for non-enzymatic polymerizations. *J. Mol. Evol.*, **12**, 91-99.

91. Kohlstaedt,L.A., Wang,J., Friedman,J.M., Rice,P.A. and Steitz,T.A. (1992) Crystal-structure at 3.5 Angstrom resolution of HIV-1 reverse-transcriptase complexed with an inhibitor. *Science*, **256**, 1783-1790.

92. Götte M., Rausch J.W., Marchand B., Sarafianos S. and Le Grice S.J. (2009) Reverse transcriptase in motion: Conformational dynamics of enzyme-substrate interactions. *Biochim. Biophys. Acta*, **1804**, 1202-1212.

93. Sarafianos,S.G., Das,K., Ding,J.P., Boyer,P.L., Hughes,S.H. and Arnold,E. (1999) Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at the polymerase active site. *Chemistry & Biology*, **6**, R137-R146.

94. Kunkel,T.A. and Bebenek,R. (2000) DNA replication fidelity. *Annual Review of Biochemistry*, **69**, 497-529.

95. Beaucage,S.L. and Caruthers,M.H. (1981) Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters*, **22**, 1859-1862.

96. Sinha,N.D., Biernat,J. and Koster,H. (1983) Beta-Cyanoethyl N,N-dialkylamino/N-morpholinomonochloro phosphoamidites, new phosphitylating agents facilitating Ease of deprotection and work-up of synthesized oligonucleotides. *Tetrahedron Letters*, **24**, 5843-5846.

97. Tener,G.M. (1961) 2-Cyanoethyl phosphate and its use in synthesis of phosphate esters. *J. Am. Chem. Soc.*, **83**, 159-168.

98. Mcbride,L.J. and Caruthers,M.H. (1983) Nucleotide Chemistry .10. An investigation of several deoxynucleoside phosphoramidites useful for synthesizing deoxyoligonucleotides. *Tetrahedron Letters*, **24**, 245-248.

99. Johnson II D.C. and Widlanski,T.S. (2004) Facile deprotection of *O*-Cbz-protected nucleosides by hydrogenolysis: an alternative to *O*-benzyl ether-protected nucleosides. *Org. Lett.*, **6**, 4643-4646.

100. Stawinski,J. and Stromberg,R. (2005) Di- and oligonucleotide synthesis using H-phosphonate chemistry. *Methods Mol. Biol*, **288**, 81-100.

101. Sobowski,M. (2000) Studies on reactions of nucleoside H-phosphonates with bifunctional reagents. Part 5. Reaction with diols. *Nucleosides, nucleotides & Nucleic Acids*, **19**, 1487-1503.

102. Hutchinson,D.W., Naylor,M. and Semple,G. (1986) Inhibition of viral nucleic acid synthesis by analogues of inorganic pyrophosphate. *Chem. Scripta*, **26**, 91-95.

103. Suo,Z.C. and Johnson,K.A. (1998) DNA secondary structure effects on DNA synthesis catalyzed by HIV-1 reverse transcriptase. *Journal of Biological Chemistry*, **273**, 27259-27267.

104. Mytelka,D.S. and Chamberlin,M.J. (1996) Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus. *Nucleic Acids Research*, **24**, 2774-2781.

105. Kang,S.M., Ohshima,K., Shimizu,M., Amirhaeri,S. and Wells,R.D. (1995) Pausing of DNA-synthesis *in-vitro* at specific loci in CTG and CGG triplet repeats from human hereditary-disease genes. *Journal of Biological Chemistry*, **270**, 27014-27021.

106. Samadashwily,G.M., Raca,G. and Mirkin,S.M. (1997) Trinucleotide repeats affect DNA replication in vivo. *Nature Genetics*, **17**, 298-304.

107. Klarmann,G.J., Schauber,C.A. and Preston,B.D. (1993) Template-directed pausing of DNA-synthesis by HIV-1 reverse-transcriptase during polymerization of HIV-1 sequences *in vitro*. *Journal of Biological Chemistry*, **268**, 9793-9802.

108. Yin,W. (2004) Structural mechanism for T7 RNA polymerase translocation. *Abstracts of Papers of the American Chemical Society*, **228**, U309.

109. Dewick,P. (2006) *Essentials of organic chemistry: for students of pharmacy, medicinal chemistry and biological chemistry*. Wiley-Blackwell, Chichester, West Sussex, England.

110. Goodman,M.F., Creighton,S., Bloom,L.B. and Petruska,J. (1993) Biochemical basis of DNA replication fidelity. *Crit. Rev. Bioch. Mol. Biol.*, **28**, 83-126.

111. Boyer,P.L., Tantillo,C., Jacobo-Molina,A., Nanni,R.G., Ding,J., Arnold,E. and Hughes,S. (1994) Sensitivity of wild-type human immunodeficiency virus type 1 reverse transcriptase to dideoxynucleotides depends on template lmength; the sensitivity of drug-resistant mutants does not. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 4882-4886.

112. Patel,P.H., Jacobo-Molina,A., Ding,J., Tantillo,C., Clark,A.D., Jr., Raag,R., Nanni,R.G., Hughes,S.H. and Arnold,E. (1995) Insights into DNA polymerization mechanisms from structure and function analysis of HIV-1 reverse transcriptase. *Biochemistry*, **34**, 5351-5363.

113. Kew,Y., Olsen,L.R., Japour,A.J. and Prasad,V.R. (1998) Insertions into the beta3-beta4 hairpin loop of HIV-1 reverse transcriptase reveal a role for fingers subdomain in processive polymerization. *J. Biol Chem.*, **273**, 7529-7537.

114. Creighton,S. and Goodman,M.F. (1995) Gel kinetic-analysis of DNA-polymerase fidelity in the presence of proofreading using bacteriophage-T4 DNA-polymerase. *Journal of Biological Chemistry*, **270**, 4759-4774.

115. Doublie,S., Tabor,S., Long,A.M., Richardson,C.C. and Ellenberger,T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature*, **391**, 251-258.

116. Esnouf,R.M. (1999) Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallographica Section D-Biological Crystallography*, **55**, 938-940.

117. Kraulis,P.J. (1991) Molscript - A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *Journal of Applied Crystallography*, **24**, 946-950.

118. Merritt,E.A. and Bacon,D.J. (1997) Raster3D: Photorealistic molecular graphics. *Macromolecular Crystallography, Pt B*, **277**, 505-524.

119. Feldmann,W. and Thilo,E. (1964) Zur chemie der kondensierten phosphate und arsenate. XXXVIII) Amidotriphosphat. *Zeit. Anorg. allg. Chem.*, **328**, 113-216.

120. Yamagata,Y., Inoue,H. and Inomata,K. (1995) Specific effect of magnesium ion on 2',3'-cyclic dAMP synthesis from adenosine and trimeta phosphate in aqueous solution. *Origin of Life and Evolution in the Biosphere*, **25**, 47-52.

121. Karsten,W.E., Ohshiro,T., Izumi,Y. and Cook,P.F. (2001) Initial velocity, spectral, and pH studies of the serine-glyoxylate aminotransferase from Hyphomicrobiuim methylovorum. *Archives of Biochemistry and Biophysics*, **388**, 267-275.

122. Liu,Y., Louie,T.M., Payne,J., Bohuslavek,J., Bolton,H. and Xun,L.Y. (2001) Identification, purification, and characterization of iminodiacetate oxidase from the EDTA-degrading bacterium BNC1. *Applied and Environmental Microbiology*, **67**, 696-701.

123. Thompson,J. and Donkersloot,J.A. (1992) N-(Carboxyalkyl)Amino Acids - Occurrence, Synthesis, and Functions. *Annual Review of Biochemistry*, **61**, 517-557.

124. Kornberg,S.R. (1956) Tripolyphosphate and trimetaphosphate in yeast extracts. *Journal of Biological Chemistry*, **218**, 23-31.

125. Eriksson,S., Munch-Petersen,B., Johansson,K. and Eklund,H. (2002) Structure and function of cellular deoxyribonucleoside kinases. *Cell Mol. Life Sci.*, **59**, 1327-1346.

126. Sabini,E., Hazra,S., Ort,S., Konrad,M. and Lavie,A. (2008) Structural basis for substrate promiscuity of dCK. *J. Mol. Biol.*, **378**, 607-621.

127. Lee,S.E., Elphick,L.M., Anderson,A.A., Bonnac,L., Child,E.S., Mann,D.J. and Gouverneur,V. (2009) Synthesis and reactivity of novel gamma-phosphate modified ATP analogues. *Bioorg. Med. Chem. Lett.*, **19**, 3804-3807.

128. Teplyakov,A., Sebastiao,P., Obmolova,G., Perrakis,A., Brush,G.S., Bessman,M.J. and Wilson,K.S. (1996) Crystal structure of bacteriophage T4 deoxynucleotide kinase with its substrates dGMP and ATP. *Embo Journal*, **15**, 3487-3497.

129. Eriksson,S., Munch-Petersen,B., Johansson,K. and Eklund,H. (2002) Structure and function of cellular deoxyribonucleoside kinases. *Cell Mol. Life Sci*, **59**, 1327-1346.

130. Sabini,E., Ort,S., Monnerjahn,C., Konrad,M. and Lavie,A. (2003) Structure of human dCK suggests strategies to improve anticancer and antiviral therapy. *Nat. Struct. Biol*, **10**, 513-519.

131. Hoard,D.E. and Ott,D.G. (1965) Conversion of mono- and oligodeoxyribonucleotides to 5'-triphosphates. *J. Am. Chem. Soc.*, **87**, 1785-1787.

132. Fields,J.H.A. and Hochachka,P.W. (1981) Purification and properties of alanopine dehydrogenase from the adductor muscle of the oyster, Crassostrea-Gigas (Mollusca, Bivalvia). *European Journal of Biochemistry*, **114**, 615-621.

133. Plaxton,W.C. and Storey,K.B. (1982) Alanopine dehydrogenase - Purification and characterization of the enzyme from Littorina-Littorea foot muscle. *Journal of Comparative Physiology*, **149**, 57-65.

134. Payne,J.W., Bolton,H., Campbell,J.A. and Xun,L.Y. (1998) Purification and characterization of EDTA monooxygenase from the EDTA-degrading bacterium BNC1. *Journal of Bacteriology*, **180**, 3823-3827.

135. Cheng,C., Fan,C., Wan,R., Tong,C., Miao,Z., Chen,J. and Zhao,Y. (2002) Phosphorylation of adenosine with trimetaphosphate under simulated prebiotic conditions. *Orig. Life Evol. Biosph.*, **32**, 219-224.

136. Fersht,A. (1985) *Enzyme Structure and Mechanism*. W.H. Freeman and Company, New York.

137. Arzumanov,A.A., Seminarov,D.G., Victorova L.S., Dyatkina,N.B. and Krayevsky,A. (1996) Gamma-phosphate-substituted 2'-deoxynucleoside 5'-triphosphate as substrates for DNA polymerase. *J. Biol. Chem.*, **271**, 24389-24394.

138. Johnson,K.A. (2010) The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochimica et biophysica acta. Proteins and proteomics*, **1804**, 1041-1048.

139. Hanes,J.W. and Johnson,K.A. (2008) Real-time measurement of pyrophosphate release kinetics. *Analytical Biochemistry*, **372**, 125-127.

140. Zlatev,I., Giraut,A., Morvan,F., Herdewijn,P. and Vasseur,J.J. (2009) Delta-di-carboxybutyl phosphoramidate of 2'-deoxycytidine-5'-monophosphate as substrate for DNA polymerization by HIV-1 reverse transcriptase. *Bioorg. Med. Chem.*, **17**, 7008-7014.